

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

Access DB# 112 888

## SEARCH REQUEST FORM

Scientific and Technical Information Center

96

Requester's Full Name Shahid Alam Examiner #: 74493 Date: 28 January 2004  
 Art Unit: 2172 Phone Number 305-2358 Serial Number 09/761,222  
 Mail Box Location: 4Y09 Results Format Preferred (circle): PAPER DISK E-MAIL

JFW

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Document Information search apparatus and method and recording medium storing document information search program therein..

Inventors (please provide full names): Seiichiro Abe

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

A search condition designating unit, a document search unit, the search conditioned designating unit transmits contents of a file designated by a user, searches similar documents from a database, forms a keyword from the file contents transmitted, . . . .

I did final rejection for this case and I think my arts are still good but I do need to get some more better art which I could not able to find. . . .

\*\*\*\*\*

## STAFF USE ONLY

## Type of Search

## Vendors and cost where applicable

Searcher: <u>Holloway</u>	Sequence (#) _____	STN _____
Searcher Phone #: <u>308-7794</u>	AA Sequence (#) _____	Dialog <u>\$ 223.99</u>
Searcher Location: <u>CPH2 YB3</u>	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>2-3-04</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>2-4-04</u>	Litigation <u>✓</u>	Lexis/Nexis _____
Searcher Prep & Review Time: <u>60</u>	Fulltext <u>✓</u>	Sequence Systems _____
Clerical Prep Time: _____	Patent Family _____	WWW/Internet <u>✓</u>
Online Time: <u>2 06</u>	Other _____	Other (specify) _____

Set	Items	Description
S1	1333203	(INDEX OR THESAURUS OR KEY) ( ) (WORD? OR TERM? OR PHRASE?) OR KEYWORD? OR KEYTERM?
S2	14276094	SEARCH? OR SEEK? OR FIND? OR QUER? OR RETRIEV? OR LOCAT?
S3	2181172	AGENT? OR IA OR SPIDER? OR CRAWLER? OR WEBCRAWLER? OR BOT - OR (SOFTWARE?) ( ) (ROBOT?) OR SOFTBOT? OR BOTS
S4	8475144	FILE? OR DOCUMENT? OR DATAFILE? OR ELECTRONIC ( ) TEXT? OR ET-EXT? OR PAGE?
S5	7139520	RELEVAN? OR RANK? OR WEIGH? OR SCORE? OR POINTS
S6	14835408	SIMILAR? OR SAME? OR CONGRUENT? OR IDENTICAL? OR CHARACTER-ISTIC? OR FEATUR?
S7	1379	S1(15N)S2(15N)S3(15N)S4
S8	168	S5(S)S6(S)S7
S9	80	RD (unique items)
S10	46	S9 NOT PY>1999
S11	46	S10 NOT PD=20010116:20030116
File	275:	Gale Group Computer DB(TM) 1983-2004/Feb 04 (c) 2004 The Gale Group
File	47:	Gale Group Magazine DB(TM) 1959-2004/Feb 02 (c) 2004 The Gale group
File	75:	TGG Management Contents(R) 86-2004/Jan W4 (c) 2004 The Gale Group
File	636:	Gale Group Newsletter DB(TM) 1987-2004/Feb 04 (c) 2004 The Gale Group
File	16:	Gale Group PROMT(R) 1990-2004/Feb 04 (c) 2004 The Gale Group
File	624:	McGraw-Hill Publications 1985-2004/Feb 03 (c) 2004 McGraw-Hill Co. Inc
File	484:	Periodical Abs Plustext 1986-2004/Jan W4 (c) 2004 ProQuest
File	613:	PR Newswire 1999-2004/Feb 04 (c) 2004 PR Newswire Association Inc
File	813:	PR Newswire 1987-1999/Apr 30 (c) 1999 PR Newswire Association Inc
File	141:	Readers Guide 1983-2004/Dec (c) 2004 The HW Wilson Co
File	239:	Mathsci 1940-2004/Mar (c) 2004 American Mathematical Society
File	553:	Wilson Bus. Abs. FullText 1982-2004/Jan (c) 2004 The HW Wilson Co
File	621:	Gale Group New Prod. Annou. (R) 1985-2004/Feb 04 (c) 2004 The Gale Group
File	674:	Computer News Fulltext 1989-2004/Jan W4 (c) 2004 IDG Communications
File	88:	Gale Group Business A.R.T.S. 1976-2004/Feb 04 (c) 2004 The Gale Group
File	160:	Gale Group PROMT(R) 1972-1989 (c) 1999 The Gale Group
File	635:	Business Dateline(R) 1985-2004/Feb 03 (c) 2004 ProQuest Info&Learning
File	15:	ABI/Inform(R) 1971-2004/Feb 03 (c) 2004 ProQuest Info&Learning
File	9:	Business & Industry(R) Jul/1994-2004/Feb 03 (c) 2004 Resp. DB Svcs.
File	13:	BAMP 2004/Jan W3 (c) 2004 Resp. DB Svcs.
File	810:	Business Wire 1986-1999/Feb 28 (c) 1999 Business Wire
File	610:	Business Wire 1999-2004/Feb 04 (c) 2004 Business Wire.
File	647:	CMP Computer Fulltext 1988-2004/Jan W4 (c) 2004 CMP Media, LLC
File	148:	Gale Group Trade & Industry DB 1976-2004/Feb 04 (c) 2004 The Gale Group

11/3,K/6 (Item 6 from file: 275)  
DIALOG(R)File 275:Gale Group Computer DB(TM)  
(c) 2004 The Gale Group. All rts. reserv.

01950763 SUPPLIER NUMBER: 18381388 (USE FORMAT 7 OR 9 FOR FULL TEXT)

The search ends: Reviewing reviewers. (Internet search sites) (Berst Mode)  
(Internet/Web/Online Service Information) (Column)

Berst, Jesse

PC Week, v13, n23, p59(1)

June 10, 1996

DOCUMENT TYPE: Column ISSN: 0740-1604 LANGUAGE: English

RECORD TYPE: Fulltext; Abstract

WORD COUNT: 580 LINE COUNT: 00048

... Oracle and Microsoft. These engines understand which topics are related to which others. They can find relevant documents even if the keyword isn't on the page .

\* Thesauri and topic searchers from companies such as Verity. These programs automatically generate and search for related terms when you enter a keyword . Inso's SearchWizard natural-language processor has similar benefits.

\* Offline researchers and search consolidators such as WebCrawler and Metacrawler, which go out to multiple search sites, then categorize and summarize what they found.

\* Collaborative filtering from companies such as Agents Inc. You tell a searcher about your preferences and it lists things that other people with similar preferences have recommended.

The victor in this race will couple a massive AltaVista-style index...

11/3,K/9 (Item 9 from file: 275)  
DIALOG(R)File 275:Gale Group Computer DB(TM)  
(c) 2004 The Gale Group. All rts. reserv.

01837736 SUPPLIER NUMBER: 17443197 (USE FORMAT 7 OR 9 FOR FULL TEXT)

As the Internet grows, so do search options. (Tippecanoe Systems Tecumseh,  
and InText CP Software Group's Retrieval Engine WebServer SDK offer  
sophisticated online information search and retrieval)

Nadile, Lisa

PC Week, v12, n37, p67(2)

Sep 18, 1995

ISSN: 0740-1604 LANGUAGE: English RECORD TYPE: Fulltext; Abstract

WORD COUNT: 508 LINE COUNT: 00045

... Retrieval Engine WebServer SDK, lets users search using naturally worded queries. The software, which offers **relevancy ranking** and summarizing **features**, includes unique HTML (Hypertext Markup Language) authoring and hyperlink technology that lets users download an HTML **document** with all links intact, officials from the San Francisco company said.

By clicking on a link in a downloaded **file**, the software automatically launches the user's browser software and connects the user to the **relevant** Web site, officials said.

The InText **Retrieval** Engine, which runs on Windows 3.1 and Unix, is priced starting at \$5,000. The product is available directly from InText.

Architext Software's **agent**-based technology, slated to debut this fall, will help users **locate relevant** information even if they are unsure of what they're looking for or the right **keywords**, according to officials from the Mountain View, Calif., company.

Architext's concept-based **searching features** will **locate documents** with **similar** content and context and generate subject groups, abstracts, and hypertext links automatically. The product's...

11/3,K/20 (Item 2 from file: 16)  
DIALOG(R) File 16:Gale Group PROMT(R)  
(c) 2004 The Gale Group. All rts. reserv.

04304609 Supplier Number: 46309387 (USE FORMAT 7 FOR FULLTEXT)  
GNN's **WEBCRAWLER INTERNET SEARCH SERVICE UPGRADE OFFERS ENHANCED  
TECHNOLOGICAL CAPABILITIES AND NEW USER INTERFACE**

PR Newswire, p416SFTU018

April 16, 1996

Language: English Record Type: Fulltext

Document Type: Newswire; Trade

Word Count: 1076

... reviewed Internet sites.

"Web surfers need ease-of-use and direction when using an Internet search engine," said Ted Leonsis, president of America Online Services. "WebCrawler's next-generation features and...

...of users exploring the Web while providing a more powerful and precise search engine."

The **WebCrawler** upgrade allows users to control how they receive search results. Users have the option to receive information by **document** title or to include **document** summary information. The title option lists search results in order of highest to lowest relevance. When summaries are requested, a **search** also returns **key phrases** which best describe the **documents** ' content. Another **search** feature of **WebCrawler** is the **Similarity Search** . Within the **search** results, **Similarity Search** , often called "more like this," assists users in **finding** additional relevant Web **pages** in their area of interest.

Significant functionality has been added to the **WebCrawler** service including the integration of GNN Select. Formerly known as WIC Select, GNN Select is...

...indexes and newsgroups as selected and reviewed by GNN's expert editorial staff. GNN Select, **located** on the front screen of the WebCrawler service, is divided into 14 distinct categories, from...

11/3,K/26 (Item 5 from file: 484)  
DIALOG(R)File 484:Periodical Abs Plustext  
(c) 2004 ProQuest. All rts. reserv.

02715774 (USE FORMAT 7 OR 9 FOR FULLTEXT)  
**Robot-generated databases on the World Wide Web**

Kimmel, Stacey

Database (DTB), v19 n1, p40-49

Feb 1996

ISSN: 0162-4105 JOURNAL CODE: DTB

DOCUMENT TYPE: Feature

LANGUAGE: English

RECORD TYPE: Fulltext; Abstract

WORD COUNT: 4074

LENGTH: Long (31+ col inches)

TEXT:

... data extraction scheme offers many of the advantages of full-text indexing. Lycos' search and **relevance ranking features** also help to account for this server's popularity. Lycos recently installed hardware and software upgrades to enhance **search** speed and availability of this service.

WebCrawler

<http://webcrawler.com>

Webcrawler was developed by Brian...

...operated by America Online, Inc. WebCrawler's database contains information on over 220,000 explored ( **retrieved** and indexed) **documents** and 3.6 million known but unexplored **documents** . While indexing, it can build its database at a rate of 1,000 **documents** per hour. WebCrawler includes HTTP, gopher, and FTP resources, and it indexes **documents** in full text, excluding stopwords.

The **WebCrawler search** form (Figure 3) allows users to select a Boolean connector (Any or All words) and set maximum **retrieval** to 10, 25, or 100 **documents** .(Figure 3 omitted) A limited truncation feature accounts for plural and singular forms of **keywords** ("s" and "es" are stripped). Items **retrieved** are **ranked** in order of **relevance** ; the results list includes the document title and its **relevance score** , displayed as a number normalized to 100. A View the Next (10, 25, 100) Results button lets users browse results beyond the maximum **retrieval** specified. The term "ebola" yielded 123 hits while "pollution" found 782 hits.

WebCrawler's easy...

11/3,K/41 (Item 2 from file: 810)  
DIALOG(R)File 810:Business Wire  
(c) 1999 Business Wire . All rts. reserv.

0910670 BW1142

**CQN INTELLISEEK: IntelliSeek Corrects and Replaces Sept. 21 Release;  
Removes Paragraph**

September 22, 1998

Byline: Business/Technology Editors

...Discovery Engine (RIDE) and the BullsEye Tracker.  
-- BullsEye Manager is the central point for launching  
**searches** . Users can save searches, analyze and refine results  
off-line, organize the information reports generated...

...bookmarks, etc. With its  
intuitive graphical interface, BullsEye Manager delivers single click  
access to favorite **searches** and the **search** history.  
-- Intelligent **Search Agents** - BullsEye features nine dynamic  
**search agents** , accessing 300+ **search** engines -- WebSearch, NewsFinder  
, PeopleTalk, BookFinder, SoftwareFinder, BusinessFinder, CollegeFinder,  
FAQFinder and HealthAnswers **Agent** . All **agents** offer an intuitive  
interface customized to the **search** context including a **query** builder  
and numerous **query** assistants (Thesaurus, Spell Checker, Sounds-like,  
etc.). BullsEye's architecture makes it easy for IntelliSeek to  
continue adding **agents** over time.

-- Rapid Information Discovery Engine (RIDE) - RIDE offers a  
set of services that underlies the operation of all **Search Agents**  
bringing state of the art information processing and linguistic  
analysis technologies under one roof to...

...include automatic  
document summarization, live highlighting and active linking of query  
keywords in the retrieved **documents** .

-- Information Tracker -- Available with BullsEye Pro only, the  
Tracker uncovers new or changed information relevant..



11/3,K/42 (Item 3 from file: 810)  
DIALOG(R)File 810:Business Wire  
(c) 1999 Business Wire . All rts. reserv.

0388675 BW040

**CAERE: Caere introductions PageKeeper Portfolio for integrated document input and management**

March 1, 1994

Byline: Business Editors & Computer Writers

...files  
and scanned or faxed documents readily accessible to individual users.  
The key features of **PageKeeper** Portfolio include the following:  
-- Automatic Index: PageKeeper Portfolio automatically indexes information without requiring any manual...

...easily integrate paper documents and electronic faxes into a database of documents.

-- Intelligent Search and **Retrieval** : PageKeeper Portfolio uses two new **search** technologies not found in other document management software packages: **Weighted Relevance Retrieval** and **Document Agent Search** . With **Weighted Relevance Retrieval** , Portfolio **searches** a database with key words and then presents those documents in order of their relevance to the **query** . **Document Agent Search** allows users to use one **document** like an **agent** to **find** other **documents** with similar or related information.

-- Data Compression: Using Caere's SuperCompression technology, **PageKeeper** Portfolio saves text and images at up to a 50:1 ratio over their uncompressed...

...This feature allows users to store graphics-intensive images as well as large text-based **documents** .

-- Visual User Interface: PageKeeper Portfolio has an intuitive graphical interface with an easy-to-use...

Set	Items	Description
S1	7991	(INDEX OR THESAURUS OR KEY) () (WORD? OR TERM? OR PHRASE?) OR KEYWORD? OR KEYTERM?
S2	1643827	SEARCH? OR SEEK? OR FIND? OR QUER? OR RETRIEV? OR LOCAT?
S3	468939	AGENT? OR IA OR SPIDER? OR CRAWLER? OR WEBCRAWLER? OR BOT - OR (SOFTWARE?) () (ROBOT?) OR SOFTBOT? OR BOTS
S4	1021094	FILE? OR DOCUMENT? OR DATAFILE? OR ELECTRONIC()TEXT? OR ET- EXT? OR PAGE?
S5	830474	RELEVAN? OR RANK? OR WEIGH? OR SCORE? OR POINTS
S6	1284597	SIMILAR? OR SAME? OR CONGRUENT? OR IDENTICAL? OR CHARACTER- ISTIC? OR FEATUR?
S7	68	S1(S)S2(S)S3(S)S4(S)S5(S)S6
S8	3	S7 AND IC=G06F-007?
S9	38	S7 AND IC=G06F?
S10	370	S2(15N)S4(15N)S1(15N)S6(15N)S5
S11	6	S10 AND IC=G06F-007?
S12	86558	S4(3N) (S5 OR S6)
S13	28	S9 AND S12
S14	33	S11 OR S13
S15	33	IDPAT (sorted in duplicate/non-duplicate order)
S16	33	IDPAT (primary/non-duplicate records only)

? show files

File 348:EUROPEAN PATENTS 1978-2004/Jan W05

(c) 2004 European Patent Office

File 349:PCT FULLTEXT 1979-2002/UB=20040129,UT=20040122

(c) 2004 WIPO/Univentio

16/5,K/2 (Item 2 from file: 349)  
DIALOG(R)File 349:PCT FULLTEXT  
(c) 2004 WIPO/Univentio. All rts. reserv.

00969519 \*\*Image available\*\*

**DYNAMIC SEARCH ENGINE AND DATABASE**

**MOTEUR DE RECHERCHE DYNAMIQUE ET BASE DE DONNEES ASSOCIEE**

Patent Applicant/Assignee:

BIOZAK INC, 50 East Saint John Street, Suite 207, San Jose, CA 95112, US,  
US (Residence), US (Nationality)

Inventor(s):

BAIDYA Ryan, 398 Boynton Avenue, #10, San Jose, CA 95112, US,  
MIFTAKHOV Valery, 50 East Saint John Street, #203, San Jose, CA 95112, US

Legal Representative:

KIM Richard C (et al) (agent), Morrison & Foerster LLP, Suite 500, 3811  
Valley Centre Drive, San Diego, CA 92130-2332, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 2002103578 A1 20021227 (WO 02103578)

Application: WO 2002US19744 20020619 (PCT/WO US0219744)

Priority Application: US 2001299708 20010619

Designated States: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU

CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP

KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO

RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG US UZ VN YU ZA ZM ZW

(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR

(OA) BF BJ CF CG CI CM GA GN GQ GW ML MR NE SN TD TG

(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZM ZW

(EA) AM AZ BY KG KZ MD RU TJ TM

Main International Patent Class: G06F-017/30

International Patent Class: **G06F-007/00** ; G06F-017/00

Publication Language: English

Filing Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 16200

**English Abstract**

An industry database (18) and method of creating same is provided. The database (18) is created in accordance with a process that includes: identifying a plurality of web sites (12) meeting at least one search criteria; automatically extracting URL addresses for each of the plurality of web sites; automatically categorizing each of the web sites and their corresponding URL addresses in accordance with a predefined category structure; and automatically indexing and storing each of the URL addresses in accordance with the predefined category structure in the database (18). A method of using a database system is also provided. The method includes: storing in a database (18), information extracted from a plurality of web sites (12), wherein the information is automatically categorized and indexed in accordance with a predefined category structure and includes a plurality of URL addresses corresponding to the plurality of web sites; receiving a user query (14); executing a search engine in response to the user query (14) that searches a subset of the stored information extracted from a subset of the plurality of web sites, and subsequently searching said subset of web sites to find additional information responsive to said user query (14).

**French Abstract**

L'invention concerne une base de donnees (18) industrielle et un procede permettant de creer cette base de donnees. Cette base de donnees (18) est

creee conformement a un processus qui consiste : a identifier une pluralite de sites Web (12) qui satisfont au moins un critere de recherche ; a extraire automatiquement des adresses URL pour chacun des sites Web ; a categoriser automatiquement chacun des sites Web et leurs adresses URL correspondantes conformement a une structure de categorie predefinie ; et a indexer et stocker automatiquement chacune des adresses URL conformement a la structure de categorie predefinie dans la base de donnees (18). L'invention concerne egalement un procede d'utilisation d'un systeme de base de donnees. Ce procede consiste : a stocker dans une base de donnees (18), des informations extraites d'une pluralite de sites Web (12), lesquelles informations sont automatiquement categorisees et indexees conformement a une structure de categorie predefinie et comprennent une pluralite d'adresses URL correspondant a la pluralite de sites Web ; a recevoir une demande d'utilisateur (14) ; pour repondre a cette demande d'utilisateur, a executer un moteur de recherche qui explore un sous-ensemble des informations stockees extraites d'un sous-ensemble de la pluralite de sites Web, puis a explorer ledit ensemble de sites Web afin d'y trouver des informations supplementaires qui repondent a la demande d'utilisateur (14).

Legal Status (Type, Date, Text)

Publication 20021227 A1 With international search report.

Publication 20021227 A1 Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

Examination 20030918 Request for preliminary examination prior to end of 19th month from priority date

International Patent Class: G06F-007/00 ...

Fulltext Availability:

Detailed Description

Detailed Description

... below the threshold value but still exceeds the minimum preset limit, the entry and all **relevant pages** are submitted to the administrator for review. Additionally, in one embodiment, changes reflecting particular types of events (e.g., new hires, new products, etc.) may be monitored using **key word search** techniques so as to alert administrators of particular changes of interest. When such changes are detected, all **relevant pages** are submitted to the administrator for review.

[0037] **Similarly**, in one embodiment, company news **pages** are periodically

scanned by the BioNews Engine for structure-changing messages, for example, I O like those describing merger or acquisition, strategic alliance etc. A set of **keywords** is defined for each such event and is matched periodically, (e.g., daily, once a week, etc.). Any other types of events may also be **searched** using appropriate **key words**. Any potentially **relevant** entries are extracted and corresponding news web **pages** and/or company names are submitted to an I 5 ...etc.) are scanned for company names present in the hifoBase database. The processing philosophy is **similar** to processing of company news pages discussed above.

[0039] In addition to the proactive auto...maximum number of companies in the biotechnology field. hi one embodiment, the engines can be **similar** to **search** engines from publicly available software such as google.com.

100741 The BioNews **search** engine provides the latest company news. In a preferred embodiment, a **search** is performed on domains (e.g., web sites) defined

4 4 9 @66

by **keywords relevant** for the news **pages** - "news", news story . news report" etc. In one embodiment, a human administrator purges the resulting list to make sure that it contains links only to head news **pages** . Alternatively or additionally, a human administrator can perform domain definition manually, determining news **page** URL addresses for each **relevant** company having a web site listed in the ...with information pertaining to potential opportunities in the industry. In one embodiment, the Opportunity Engine **searches** pre-selected resources for **relevant** information.

Such resources may include, for example, specific pages of university web sites, government research...Technologies Ltd., Elan Corporation PLC, Ethypharm, etc.

100761 In a preferred embodiment, information is **retrieved** and updated from these pre-selected web pages in accordance with the methods discussed above.

Additionally, the **retrieved** information may be automatically classified, indexed and stored in the InfoBase in a **similar** fashion to the techniques discussed above.

[0077] In one embodiment, the Opportunity Engine **searches** indexed web **pages** having URLs and corresponding content stored in the InfoBase, when such web **pages** satisfy user criteria (e.g., all web **pages** associated with diagnostic companies). As described above, potentially **relevant pages** may be identified using **key word** and/or class field **searches** (e.g., "licens\* and diagnostic") entered by a member/user. Opportunity information/content stored in the InfoBase may be updated in a **similar** fashion to the techniques described above for updating BioField and BioNews information.

[00781 In...

16/5,K/14 (Item 14 from file: 349)  
DIALOG(R)File 349:PCT FULLTEXT  
(c) 2004 WIPO/Univentio. All rts. reserv.

00794300 \*\*Image available\*\*

**INDEXING A NETWORK WITH AGENTS**

**INDEXAGE D'UN RESEAU AU MOYEN D'AGENTS**

Patent Applicant/Assignee:

360 POWERED CORPORATION, Suite 300, 190 Queen Anne, Seattle, WA 98109, US  
, US (Residence), US (Nationality)

Inventor(s):

TRIPP Gary W, 9836 Miller Road, Bainbridge Island, WA 98110, US,  
MEADWAY Michael D, 1803-129th Place Southeast, Snohomish, WA 98110, US,  
DUGUAY Claude E, 2835 Boyer Avenue East, Seattle, WA 98102, US,

Legal Representative:

HALEY Jeffrey T (et al) (agent), Graybeal Jackson Haley LLP, Suite 350,  
155-108th Avenue N.E., Bellevue, WA 98004-5901, US,

Patent and Priority Information (Country, Number, Date):

Patent: WO 200127793 A2-A3 20010419 (WO 0127793)

Application: WO 2000US21020 20000801 (PCT/WO US2000021020)

Priority Application: US 99419405 19991014; US 2000575974 20000523

Designated States: AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK

DM EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR

LS LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ

TM TR TT TZ UA UG UZ VN YU ZA ZW

(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE

(OA) BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG

(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW

(EA) AM AZ BY KG KZ MD RU TJ TM

Main International Patent Class: **G06F-017/30**

Publication Language: English

Filing Language: English

Fulltext Availability:

Detailed Description

Claims

Fulltext Word Count: 19700

**English Abstract**

A search engine utilizes a bottom-up approach to index the content of a network with agent programs running on each source computer instead of relying on a top-down approach as used by conventional search engines. The network being indexed may be any network, including the global computer network or an intranet. Instead of using a central site including spidering software to recursively search all linked web pages and generate a search index of the Internet, independent distributed components or agent programs are located at each web site and report meta data about objects at the web site to the central server. A central catalog of object references is compiled on the central site or sites from the meta data reported from each web site. One or more brochure files may also be created and stored on each web site to provide conceptual or non-key-word data about the site, such as target demographics and categorization information. This conceptual information is then utilized in constructing the central catalog so that more accurate search results may be generated for search queries applied to the catalog.

**French Abstract**

Selon l'invention, un moteur de recherche met en oeuvre une approche base/sommet pour indexer le contenu d'un reseau, au moyen de programmes comprenant des agents et fonctionnant sur chaque ordinateur source, plutot que de mettre en oeuvre une approche sommet/base, comme c'est le

cas pour les moteurs de recherche classiques. Le reseau a indexer peut etre n'importe quel reseau, notamment le reseau informatique mondial ou un Intranet. Au lieu d'utiliser un site central comprenant un logiciel de balayage du Web, pour rechercher de maniere recursive toutes les pages du Web liees et produire un index de recherche de l'Internet, des composants repartis, independants, ou programmes comprenant des agents, sont situes au niveau de chaque site Web et signalent au serveur central les metadonnees relatives a des objets situes au niveau du site Web. Un catalogue central de references d'objets est compile sur le ou les sites centraux, a partir des metadonnees signalees depuis chaque site Web. Un ou plusieurs fichiers de brochures peuvent egalement etre crees et stockes sur chaque site Web, afin de constituer des donnees conceptuelles ou de mots non cles relatifs au site, telles que des informations demographiques cibles et des informations de categorisation. Ces informations conceptuelles sont utilisees ensuite dans la construction du catalogue central, de facon que des resultats de recherche plus precis puissent etre produits pour des demandes de recherches effectuees aupres du catalogue.

Legal Status (Type, Date, Text)

Publication 20010419 A2 Without international search report and to be republished upon receipt of that report.

Examination 20010809 Request for preliminary examination prior to end of 19th month from priority date

Search Rpt 20040108 Late publication of international search report

Republication 20040108 A3 With international search report.

Republication 20040108 A3 Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

Main International Patent Class: **G06F-017/30**

Fulltext Availability:

Detailed Description  
Claims

Detailed Description

... The indexer also can retrieve the contents of an html page to extract 1 5 **relevant document** information and index the document so that subsequent search queries may be applied on indexed...

...allows visitors to the web server to apply search queries, and returns a list of **documents ranked** by confidence in response to the search queries. Since the program resides on the web...brochure

17. Link URL Link Table  
18. Html tag information  
19. XML tag information  
26

. **Ranking**

Table 5- **Agent** Created Products Catalog

1. iii. Type of product  
2. Category three letters representing General, Specific...

...URL,

6. Unique Record Identifier  
7. iv. Product Number  
8. V. Product price  
9. A. **Feature** or option  
10. **Feature** or option  
11. **Feature** or option  
12. Link URL Link Table

Table 6- **Agent** Created Articles & **Documents** Table

I vii. Type of Articles or **Documents**

2. Category - three letters representing General, Specific, and Special Interest Categories

...1, 2, 3, 4, 5, 6, 7, 8, 9 & 10

4. Subject of Articles or **Documents**

5. Site URL,

6. Unique Record Identifier

7. viii. Date

8. ix. Author

9. X. Source of Articles or **Documents**

10.

11.

27

. Link URL Link Table

Table 7-Agent Created MP3 Table Fields

I...to classify documents as they are found, and to assign concepts and the concept of **relevancy** strength to each **document** during parsing. The agent would 1 5 thereafter store these concepts as standard name/value extraction, where the unique statistical and logical **characteristics** of image **files** processed 0 by the agent are determined and forwarded to a central site for later...

...understood by those skilled in the art.

During operation, the agent can parse local image **files** to extract "**features**" contained within the images. For example, a file containing a picture of a face can...an example of this type of AI analysis of local files by the agent.

The **agent** may also be used to determine the relative importance of a 1 5 **document** as a source or reference of information stored in linked **documents**. As an example of adult site detection, the **agent** might use a database consisting of a list of the Companies have developed **search** engine technologies that **search** based upon pattern matching and content **weighting** techniques. For example, IBM has developed Query By Image Content ("QBIC") and a system known...

...below. The QBIC and CLEVER systems would be capable of using data produced by the **agent** for image, audio, and link information. The QBIC system uses a pattern-matching engine embedded into an IBM DB2 database system to compare image **characteristics** against a sample image. The results of such comparisons are then **retrievable** via a Structured Query Language ("SQL") statement. The QBIC system is intended for use in a **keyword** environment, where a **keyword search** produces an initial set of images which are then used as comparison templates and compared against the patternmatching engine. The CLEVER system determines information source **documents** or "hubs"

41

from URLs collected from one or more web sites. This is **similar** in concept to the methods described this year in a Scientific American article, but the CLEVER system is actually running. A source **document** is one that is referenced by many web **pages** or URLs, sometimes several levels removed from the **document** itself. A hub is defined as a **page** containing a series of links to other sites or source **documents**, and is often referred to as a "links" **page**.

In both the QBIC and CLEVER systems, a source index or collection of information is...applied to any system which requires transformation of source data into a series of data **points**. A sound **file**, for example, can be represented either as the time-series data (the actual digitized sound...



16/5,K/18 (Item 18 from file: 349)  
DIALOG(R) File 349:PCT FULLTEXT  
(c) 2004 WIPO/Univentio. All rts. reserv.

00764264 \*\*Image available\*\*

**SYSTEM AND METHOD FOR DOCUMENT MANAGEMENT BASED ON A PLURALITY OF KNOWLEDGE  
TAXONOMIES**

**SYSTEME ET PROCEDE DE GESTION DE DOCUMENTS BASES SUR PLUSIEURS TAXONOMIES  
DES CONNAISSANCES**

Patent Applicant/Assignee:

KANISA INC, 1595 Kingswood Drive, Hillsborough, CA 94010, US, US  
(Residence), US (Nationality)

Inventor(s):

ANGEL Mark, 20332 Pinntage Park, Cupertino, CA 95014, US  
BILINSKI Alan, 13 County Fair, St. Louis, MO 63141, US  
COPPERMAN Max, 233 Sunset Avenue, Santa Cruz, CA 95060, US  
FRATKINA Raya, 673 Royston Lane #236, Hayward, CA 94544, US  
HUFFMAN Scott B, 195 Opal Avenue, Redwood City, CA 94062, US  
KAY David Beesley, 18275 Knutul Road, Los Gatos, CA 95033, US  
MOTWANI Rajeev, 251 Stanford Avenue, Palo Alto, CA 94306, US  
PETERS Stanley, 128 Hillside Avenue, Menlo Park, CA 94025, US  
PRUETT Rodney, 772 Coronado Lane, Foster City, CA 94404, US  
RUDY Jeffrey H, 1074 Foxhurst Way, San Jose, CA 95120, US

Legal Representative:

GARRETT Arthur S, Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.,  
1300 I Street, N.W., Washington, DC 20005-3315, US

Patent and Priority Information (Country, Number, Date):

Patent: WO 200077690 A1 20001221 (WO 0077690)  
Application: WO 2000US16444 20000615 (PCT/WO US0016444)  
Priority Application: US 99139509 19990615

Designated States: AE AG AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE  
DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC  
LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI  
SK SL TJ TM TR TT TZ UA UG UZ VN YU ZA ZW  
(EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE  
(OA) BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG  
(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW  
(EA) AM AZ BY KG KZ MD RU TJ TM

Main International Patent Class: **G06F-017/30**

Publication Language: English

Filing Language: English

Fulltext Availability:

Detailed Description  
Claims

Fulltext Word Count: 31064

**English Abstract**

A method and system are disclosed for organizing and retrieving information through the use of taxonomies. Documents stored in the organization and retrieval subsystem may be manually or automatically classified into a predetermined number of taxonomies. In operation, automatic term extractor creates a list of terms that are indicative of the subject matter contained in the documents. A term analysis system assigns the relevant terms to one or more taxonomies, and a suitable algorithm is then used to determine the relatedness between each list of terms and its associated taxonomy. The system then clusters documents for each taxonomy in accordance with the weights ascribed to the terms in the taxonomy's list and a directed acyclic graph (DAG) hierarchical structure is created. The present invention may then be used to aid a researcher or user in quickly identifying **relevant documents**, in response to an inputted query.

#### French Abstract

L'invention concerne un procede et un systeme qui utilisent des taxonomies pour organiser et extraire des informations. Des documents stockes dans les sous-systemes d'organisation et d'extraction peuvent etre classes manuellement ou automatiquement en un nombre preetabli de taxonomies. Pendant le fonctionnement, un extracteur terminologique automatique etablit une liste de termes indiquant la matiere traitee dans les documents. Un systeme d'analyse terminologique attribue les termes pertinents a une ou plusieurs taxonomies, et un algorithme approprie est ensuite utilise pour determiner le rapprochement entre chaque liste de termes et la taxonomie qui lui est associee. Le systeme regroupe ensuite les documents pour chaque taxonomie, conformement aux poids attribue aux termes figurant dans la liste de la taxonomie, aux fins de creer un graphe acyclique oriente (DAG) ou une structure hierarchique. Le procede de l'invention peut ainsi etre utilise pour aider un chercheur ou un utilisateur a identifier rapidement des documents pertinents en reponse a une demande entree.

Legal Status (Type, Date, Text)

Publication 20001221 A1 With international search report.

Publication 20001221 A1 Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.

Examination 20010322 Request for preliminary examination prior to end of 19th month from priority date

Main International Patent Class: G06F-017/30

Fulltext Availability:

Claims

#### English Abstract

...present invention may then be used to aid a researcher or user in quickly identifying **relevant documents**, in response to an inputted query.

#### Claim

... present invention may then be used to aid a researcher or user in quickly identifying **relevant documents**, in response to an inputted query. It may be appreciated that both a document's...

...relevant response than a contentbased retrieval, which is driven by the actual words in the **document**. Additional **features** and advantages of the invention will be set forth in the ...the word "truck" was in the query. Documents tagged to "Trucks" are likely to be **relevant**. **Documents** tagged to the child concept node "Pick-up" may or may not be relevant, but...control the process of taxonomy tag identification 15 using the text classifiers. These include threshold **scores** for tagging either **documentknowledge** containers or question-knowledge containers, and maximum numbers of tags to assign from each topic...may also consider sentence boundaries, section boundaries or page boundaries are considered as possible slice **points**. In general, a **document** should be sliced at points where there is a fairly substantial and permanent shift in...

...and paragraphs N-2 and N-3, etc, up to some window size W; and **similarly** between 10 paragraph N and N+ 1; and between N- I and N+ 1, N...

...and slices 820. As shown in FIG. 8, the slicing algorithm has split the

example **document** into 6 **similarly** -sized slices 820 a-f. Each slice 820 contains 1-3 paragraphs 720, and 2...tag taxo=Industry tagid=fgl weight=1.0 attribution--human>Federal Government<tag>  
 <tag taxo= **Document** -Source tagid=reutl **weight** --1.0 attribution=human>Reuters</tag>  
 <taxonomy-tags>  
 </context>  
 <content>  
 IRS Reform Bill Passes  
 Dateline...be identified by the text classifiers might be:  
 Government Agencies: Federal: Legislative: Congress with (estimated) **weight** = 0.65  
 Government Agencies: Federal: Executive: IRS with (estimated) **weight** = 0.75; and  
 Government Issue: Legislation: New Legislation with (estimated) **weight** = 0  
 Each of these three tags have associated terminology that evidences the presence of the...

```
...time>09:3 6:00</time>
</submission-time>
<taxonomy-tags>
<tag taxo=Industry tagid=fgl weight =1.0 attribution=human>Federal
Government</tag>
<tag taxo= Document -Source tagid=reutl weight =1.0
attribution=human>Reuters</tag>
<tag taxo=Government-Agencies tagid=conl weight =0.65
attribution=machine>Congress</tag>
<tag taxo=Government-Agencies tagid=irs I weight =0. 75
attribution=machine>IRS</tag>
<tag taxo=Government-Issues tagid=nll weight =0.50
attribution=machine>New Legislation<tag>
</taxonomy-tags>
</context>
<content>
<title><evid value=high...
```

```
...year><time>09:36:00</time>
</submission-time>
<taxonomy-tags>
<tag taxo=Industry tagid=fgl weight =1.0 attribution--human>Federal
Government</tag>
<tag taxo= Document -Source tagid=reutl weight =1.0
attribution--human>Reuters</tag>
<tag taxo=Government-Agencies tagid=conl weight =0.65
attribution=machine>Congress</tag>
<tag taxo=Goveniment-Agencies tagid=irs I weight =0.75
attribution=machine>IRS</tag>
<tag taxo=Government-Issues tagid=nll weight =0.50
attribution=machine>New Legislation</tag>
<tag taxo=Government-Officials tagid=lottl
attribution=lexical...
```

...Government Agencies" taxonomy as a topic taxonomy rather than a lexical one. Therefore, tagging this **document** to, e.g., "IRS" was done using a text-classifier over the entire text to...  
 ...directly to the concept-node "IRS". The topic taxonomy for Government Agencies indicates that the **document** concerns the tagged agencies; a lexical taxonomy that both can be useful for **retrieving documents** .

The next step in the process involves using symbolic rules and reasoning in order to refine the set of tags applied to the **document** . For example, the output of this process may be the determination that another concept node that might be **relevant** to our example content is:

Government Issues:Legislation:Tax Legislation

15 A knowledge-based transformation that might infer the **relevance** of this concept node is:

If content is tagged to Govern-ment Agencies:Federal:Executive:IRS with **weight** above 0.60 and content is tagged to any node under Government Agencies:Government Issues:Legislation with **weight** X where X is greater than 0.35, add tag Government Issues:Legislation:Tax Legislation to the content with **weight** X. Finally, the system stores the results as a knowledge container in its data store. If the **document** had been longer, the system could optionally invoke slicing to break the **document** into multiple, contiguous sections with different topics assigned to each section. In this case, however...

...this description will address a process for creating a knowledge map from a collection of **documents** . As explained above, 5taxonomies, and by extension knowledge maps, may be manually constructed based on...

...quality of operation. The input into the knowledge map generation mechanism is a set of **documents** and a set of "target" taxonomy root nodes. The output is a knowledge map. A...

...starting point for knowledge map generation, as shown in FIG. 9, is the collection of **documents** that will be managed by the e-Service Portal (step 902). This collection will 20...

...to be placed in knowledge containers. In one embodiment, the generation corpus has the following **characteristics** : (1) the **documents** in the corpus are a statistically valid sample of the **documents** to be managed; (2) there are at least 1,000 and less than 30,000 **documents** ; (3) there are at least the equivalent of 500 **pages** of text and no more than 50,000 **pages** of text; and (4) the **documents** are decomposable into ASCII text. The knowledge map generation process described below is language 5 independent. That is, so long as the **documents** can be converted into **electronic text** , the process is also independent of **document** format and type. The second input into the process (step 904) is a set of...

...a valid input. First, the concept-nodes do not overlap. Second, the concept-nodes are **relevant** . 15 Third, the concept-nodes are orthogonal. The purpose of each root concept-node is...

...in the knowledge map. Overlap occurs when two root nodes are provided that are actually **identical** or nearly 20 **identical** . In effect, the root concept-nodes are synonyms, and taxonomies generated from them would cover substantially the **same** portion and aspect of the knowledge domain. For example, the root nodes "Geography - The World..."

...are ascribed to a particular root, then that root concept-node is probably not 10 **relevant** . The cure is to eliminate the concept-node from the input set and to re...

...is to have one taxonomy for each orthogonal view of knowledge within the corpus.

Each **document** may have one or more taxonomy tags into each taxonomy. In

an orthogonal knowledge map...of knowledge map generation. If there is little or no cross-tagging between two taxonomies ( **documents** tagged to one taxonomy are not tagged to another taxonomy), then non-orthogonality can be...

...level concept node and to re-initiate the knowledge map generation mechanism. Assuming valid inputs ( **documents** and root concept-node set), the invention will produce a valid output. As stated earlier...

...node in the input set. As shown in FIG. 9, the first step (904) is **document** collection. The generation corpus is a representative sample of **documents** from a single coherent knowledge domain, the representation of which meets the needs of a specific business problem or domain. In one typical scenario, an enterprise has a corpus of **documents** over which they would like to provide the **retrieval** and display capabilities described earlier in this specification. In that case, the generation corpus would be a subset of the enterprise's corpus of **documents**. The subset may be manually identified. In another scenario, the knowledge domain is well-defined...

...the enterprise does not yet have a corpus covering the domain. In this case, representative **documents** must be found and accumulated to form the generation corpus If the available corpus is...

...the maximum size prescribed above, sampling procedures may be employed to choose a subset of **documents** for use in the generation corpus. As shown in step 906, the next step is to convert the **documents** into XML marked text as described above in the portion of the **document** that addressed autocontextualization. Next, in step 908, the system performs root concept-node collection and...

...common to all root concept-nodes within the knowledge map). In a preferred embodiment, a **file** is prepared designating the set of root concept-nodes. This **file** is provided as an input to knowledge map generation and includes one record (with all...

...910, the system identifies and inputs the generation corpus. In one embodiment, a **file** listing each individual 10 **document** in the generation corpus and its physical **location**, one per line, is provided as an input to knowledge map generation. In step 912, term extraction is then performed. Using any valid algorithm for term **feature** extraction, a list of corpus terms is generated. The term list is ordered by frequency or **weight**. This term list includes all indicators of meaning within the generation corpus. The term list is a function of the generation corpus **documents** - the text of these **documents** is read and parsed to produce the list. A term may have any (or none) of the following **characteristics** in any combination: a term may be case-sensitive ( the term "jaguar" is distinct from...

...the knowledge domain associated with the generation corpus. The SME designates whether the term is **relevant** to each of the taxonomies in the input set. Each term may be **relevant** in zero to N taxonomies where N is the number of root concept nodes. For example, the term "jaguar" may be **relevant** to the taxonomy on "Mammals" and the taxonomy on "Automobiles". The result of this step...

...each root concept node. The terms extracted in step 912 are automatically provisionally designated as **relevant** to zero or more taxonomies according to their **similarity** to the SME-generated term

sets, using any word- **similarity** measures or algorithms from the fields of computational linguistics and information **retrieval** . These designations are presented to the SME for validation. Next, in step 916, the system...

- ...every other taxonomy; and (3) a list of terms assigned to each taxonomy ordered by **weight** or frequency. Processing then flows to step 920, where the system performs diagnosis for irrelevant...
- ...system determines whether any taxonomy is assigned a small number or percentage of the term/ **features** . If there are taxonomies that are assigned to a small number of terms/ **features** , processing flows to step 924 and the concept node is removed from the input list...determines that there is not overlap or nonorthogonality, processing flows to step 934, where term **weighting** is performed. Using 10 any standard algorithm for **weighting** a list of **features** in terms of relative importance, the term list for each taxonomy is **weighted** . Terms have a unique **weight** in relationship to each taxonomy to which they are ascribed. So, the term "jaguar" may have a low **weight** in relationship to the "Mammal" taxonomy and a high **weight** in relationship to the "Automobile" taxonomy and a zero **weight** (non-ascribed) in 15 relationship to a "Geography" taxonomy. Optionall y, the system may in step 936, subject the term **weights** generated in step 934 to review by an SME. The SME may then enter a new **weight** , replacing the computer-generated **weight** . One **weighting** algorithm has the following key **characteristics** :

1 Terms with a high weight in one taxonomy have suppressed weights in all other...based on vocabulary usage. That is, words occurring in a query that appear with the **same** frequency in every **document** contribute nothing to the **rank** of any **document** . At the other end of the spectrum, a query word that appears in only one document, and occurs many times in that **document** , greatly increases the **rank** of that **document** . **Ranking** takes into account the occurrences of a word both in the **document** being **ranked** and in the collection at large --- to be precise, in the indexed collection. To be...

- ...of words that a search engine takes into account. The mathematical expression commonly associated with **ranking** is:

**Document Rank** =  $Tf / df$

where, Tf = number of times a term occurs in a document

df = document...green

1510

lffi@ - blue

FIGn 18

Ran g: earc Eng ne F

20

For each **document** 0.

the **rank** returned by

the search engine is 82adjusted by ... 818079

Forpurposes of illustration,

rank is shown...and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC, IBM-TDB

C. **DOCUMENTS** CONSIDERED TO BE **RELEVANT**

Category o Citation of document, with indication, where appropriate, of the relevant passages Relevant to...

- ...not cited to understand the principle or theory underlying the considered to be of particular **relevance** invention 8E1 earlier **document** but published on or after the international W **document** of particular **relevance** ; the claimed invention

filing date cannot be considered novel or cannot be considered to  
81...

...is taken alone which is cited to establish the publication date of  
another o'Yo **document** of particular **relevance** ; the claimed invention  
citation or other special reason (as specified) cannot be considered to  
involve...

...the international filing date but in the art. later than the priority  
date claimed o&' **document** member of the **same** patent family Date of  
the actual completion of the international search Date of mailing of...

...1 of 3

INTERNATIONAL SEARCH REPORT Into onal Application No

PCTAS 00/16444

C.(Continuation) **DOCUMENTS** CONSIDERED TO BE **RELEVANT**

Category Citation of **document** , with indication, where appropriate, of  
the relevant passages Relevant to claim No.

X WONG J...2 of 3

INTERNATIONAL SEARCH REPORT Into onal Application No

PCTAS 00/16444

C.(Continuation) **DOCUMENTS** CONSIDERED TO BE **RELEVANT**

Category Citation of **document** , with indication, where appropriate, of  
the relevant passages Relevant to claim No.

A CHAKRABARTI S...

Set	Items	Description
S1	9991	(INDEX OR THESAURUS OR KEY) ( ) (WORD? OR TERM? OR PHRASE?) OR KEYWORD? OR KEYTERM?
S2	1152186	SEARCH? OR SEEK? OR FIND? OR QUER? OR RETRIEV? OR LOCAT?
S3	1006393	AGENT? OR IA OR SPIDER? OR CRAWLER? OR WEBCRAWLER? OR BOT - OR (SOFTWARE?) ( ) (ROBOT?) OR SOFTBOT? OR BOTS
S4	283095	FILE? OR DOCUMENT? OR DATAFILE? OR ELECTRONIC ( ) TEXT? OR ET- EXT? OR PAGE?
S5	1033005	RELEVAN? OR RANK? OR WEIGH? OR SCORE? OR POINTS
S6	2515527	SIMILAR? OR SAME? OR CONGRUENT? OR IDENTICAL? OR CHARACTER- ISTIC? OR FEATUR?
S7	21	S1 AND S2 AND S3 AND S4
S8	8	S7 AND (S5 OR S6)
S9	9	S7 AND IC=(G06F-015? OR G06F-007?)
S10	14	S8 OR S9
S11	14	IDPAT (sorted in duplicate/non-duplicate order)
S12	13	IDPAT (primary/non-duplicate records only)
S13	70	S2 AND S4 AND S5 AND S6 AND S1
S14	69	S13 NOT (S7 OR S8 OR S9)
S15	6	S14 AND IC=G06F-007?
S16	69	S14 AND IC=G06F?
S17	55	S16 NOT AD>20010116
S18	13928	S4(5N) (S5 OR S6)
S19	33	S17 AND S18

? show files

File 347:JAPIO Oct 1976-2003/Sep(Updated 040105)

(c) 2004 JPO & JAPIO

File 350:Derwent WPIX 1963-2004/UD,UM &UP=200408

(c) 2004 Thomson Derwent



19/5/1 (Item 1 from file: 347)  
DIALOG(R) File 347:JAPIO  
(c) 2004 JPO & JAPIO. All rts. reserv.

07328607 \*\*Image available\*\*

**KEYWORD** EXTRACTING DEVICE AND INFORMATION **RETRIEVING** DEVICE

PUB. NO.: 2002-197095 [JP 2002197095 A]  
PUBLISHED: July 12, 2002 (20020712)  
INVENTOR(s): UMEMURA KYOJI  
TAKENAMI YOSHINORI  
KISHIDA MASAHIRO  
APPLICANT(s): SUMITOMO ELECTRIC IND LTD  
UMEMURA KYOJI  
APPL. NO.: 2000-394194 [JP 2000394194]  
FILED: December 26, 2000 (20001226)  
INTL CLASS: **G06F-017/30**

#### ABSTRACT

PROBLEM TO BE SOLVED: To extract a **keyword** from a **document** without necessity of a dictionary.

SOLUTION: A **keyword** extracting device includes a suffix **file** generating part 22 to receive a group of **documents** and to generate a suffix **file** to be described later from the group of **documents**, a suffix **file** storage part 24 to store the suffix **file**, a punctuating part 28 to receive an optional **document** to be included in the group of **documents** or a **document** in the same field as the group of **documents** and to punctuate the **document** at a break of a sentence such as punctuation marks, a **score** calculating part 26 to properly punctuate the sentence based on the suffix **file** and the sentence supplied from the punctuating part 28 and to calculate appearance frequency a, a degree B of concentration of appearance and **weight**, etc., to be described later, an operation result storage part 30 to store an operation result, a **document** separating part 32 to punctuate the **document** into candidates of the **keyword** based on the operation result and a narrowing part 34 to narrow down the candidates of the **keyword**.

COPYRIGHT: (C)2002,JPO

19/5/3 (Item 3 from file: 347)  
DIALOG(R)File 347:JAPIO  
(c) 2004 JPO & JAPIO. All rts. reserv.

07146620 \*\*Image available\*\*

**SIMILAR**      **DOCUMENT**      **RETRIEVAL**      DEVICE AND RELATIVE      **KEYWORD**      EXTRACT  
DEVICE

PUB. NO.:      2002-014999 [JP 2002014999 A]  
PUBLISHED:      January 18, 2002 (20020118)  
INVENTOR(s):      SUGANO YUJI  
APPLICANT(s):      MATSUSHITA ELECTRIC IND CO LTD  
APPL. NO.:      2000-195075 [JP 2000195075]  
FILED:      June 28, 2000 (20000628)  
INTL CLASS:      **G06F-017/30**

#### ABSTRACT

PROBLEM TO BE SOLVED: To attain **retrieval** of similar documents and also extraction of relative **keywords** with high accuracy and robustness by analyzing interpedently the **weighted** main components at both **document** and **keyword** sides in accordance with the **keyword** appearance frequency and obtaining a **feature** vector.

SOLUTION: Three types of data are produced on **keyword** appearance frequency 103, **document** length 105 and **keyword** weight 107 respectively. Then profile vectors 111 and 109 of **documents** and **keywords** are calculated and the **weighted** main component analyses 112 and 114 are carried out independently of each other in consideration of the length 105 and **keyword** weight 107 for obtaining the **feature** vectors of each **document** and **keyword** . Then a **document** and a **keyword** having high **similarity** to the **feature** vector that is calculated from the **retrieval** /extraction condition are obtained and displayed.

19/5/7 (Item 7 from file: 347)  
DIALOG(R)File 347:JAPIO  
(c) 2004 JPO & JAPIO. All rts. reserv.

06083594 \*\*Image available\*\*  
AUTOMATIC EXTRACTION DEVICE FOR RELATIVE **KEYWORD** , **DOCUMENT** **RETRIEVING**  
DEVICE AND **DOCUMENT** **RETRIEVING** SYSTEM USING THESE DEVICES

PUB. NO.: 11-025108 [JP 11025108 A]  
PUBLISHED: January 29, 1999 (19990129)  
INVENTOR(s): SATO MITSUHIRO  
NOGUCHI NAOHIKO  
SUGANO YUJI  
NOMOTO MASAKO  
INABA MITSUAKI  
FUKUSHIGE TAKAO  
APPLICANT(s): MATSUSHITA ELECTRIC IND CO LTD  
APPL. NO.: 09-176822 [JP 97176822]  
FILED: July 02, 1997 (19970702)  
INTL CLASS: **G06F-017/30**

#### ABSTRACT

PROBLEM TO BE SOLVED: To automatically extract a relative **keyword** which is matched with the **characteristics** of a **document** to be practically **retrieved** and which is capable of obtaining one or more **retrieval** results at the time of executing **retrieval** using the **keyword** .

SOLUTION: An automatic extraction device for relative **keywords** is provided with a **document** set selection part 19 for specifying a partial set of each **document** based on the attribute information, input **retrieval** expression, etc., of the **document** , a word statistic information management part 17 for managing the statistic information of respective words in the whole objective **document** 11 and words appearing in each **document** as well as their statistic information 15; and a word **ranking** part 18 for calculating the importance of each word appearing in a partial set of a certain **document** and for aligning respective words in the order of importance, wherein the management part 17 quickly **finds** out the statistic information of respective words in the whole **document** and a specified partial set of the **document** . Consequently, words appearing in a certain **document** set can be **ranked** based on their importance and a part of the **ranked** words can be presented as a relative **keyword** .

COPYRIGHT: (C)1999,JPO

19/5/11 (Item 11 from file: 347)  
DIALOG(R)File 347:JAPIO  
(c) 2004 JPO & JAPIO. All rts. reserv.

04751967 \*\*Image available\*\*  
DOCUMENT RETRIEVAL DEVICE

PUB. NO.: 07-044567 [JP 7044567 A]  
PUBLISHED: February 14, 1995 (19950214)  
INVENTOR(s): SATO OSAMU  
APPLICANT(s): FUJITSU LTD [000522] (A Japanese Company or Corporation), JP  
(Japan)  
APPL. NO.: 05-188243 [JP 93188243]  
FILED: July 29, 1993 (19930729)  
INTL CLASS: [6] G06F-017/30  
JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications); 45.2  
(INFORMATION PROCESSING -- Memory Units)

#### ABSTRACT

PURPOSE: To provide a document retrieval device capable of obtaining an absolutely sufficient retrieved result with the retrieval of one time by retrieving similar documents from a document data base with the document itself as a retrieval key.

CONSTITUTION: This document retrieval device is constituted of a retrieval key word set generation means 2 for analyzing an input document 1 and generating a retrieval key word set 3 for which weighing corresponding to document component elements is performed and a document retrieval means for retrieving the document data base based on the retrieval key word set 3, calculating the weight of respective matched key words for each document obtained as a result and obtaining cumulative weight for the document of the retrieved result. Since the cumulative weight indicating the degree of similarity with the input document is added to the retrieved result, a user can efficiently select the retrieved result by referring to it.

19/5/22 (Item 4 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2004 Thomson Derwent. All rts. reserv.

013978328 \*\*Image available\*\*  
WPI Acc No: 2001-462542/200150

**Keyfact-based text search system, apparatus and method for searching  
keyfact-based text therewith**

Patent Assignee: KOREA ELECTRONICS & TELECOM RES INST (KOEL-N); ELECTRONICS  
& TELECOM RES INST (ELTE-N)

Inventor: JANG M G; JUN M S; JUNG G T; PARK S Y; CHONG K T; JANG M; JUN M

Number of Countries: 002 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
KR 2001004404	A	20010115	KR 9925035	A	19990628	200150 B
US 6366908	B1	20020402	US 99475743	A	19991230	200226

Priority Applications (No Type Date): KR 9925035 A 19990628

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
KR 2001004404	A		1	G06F-017/30	
US 6366908	B1			G06F-017/30	

Abstract (Basic): KR 2001004404 A

NOVELTY - A keyfact-based text **search** system is provided to display concepts of a **document** with a couple of an object and property and to index and **search** text based on the couple-displayed data.

DETAILED DESCRIPTION - In a keyfact-based text **search** system, a keyfact sampling device(11) samples keyfacts from plural **key words** having the improved vagueness in speech by analyzing a **document** group to be **searched** and question of a user. A keyfact index device(12) saves a keyfact list of the entire **document** groups in a **search** structure of keyfact as well as calculates frequency of various keyfacts in the **document** group to be **searched**. A keyfact **search** device(13) receives the key facts about the question of the user and the other ones of the **document** group. The keyfact **search** device defines a keyfact-based **search** model and outputs the **similar document** to the question by considering a **weighting** constant depending on the type of keyfacts.

pp; 1 DwgNo 1/10

Title Terms: BASED; TEXT; **SEARCH** ; SYSTEM; APPARATUS; METHOD; **SEARCH** ;  
BASED; TEXT

Derwent Class: T01

International Patent Class (Main): G06F-017/30

File Segment: EPI

19/5/23 (Item 5 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2004 Thomson Derwent. All rts. reserv.

013963149 \*\*Image available\*\*

WPI Acc No: 2001-447363/200148

XRPX Acc No: N01-331008

Document **screening based on similarity of contents intrinsic to**  
**search document , determines frequency of keywords in each of**  
**screened documents preclassified subjectwise**

Patent Assignee: TOSHIBA COMPUTER ENG KK (TOSH-N); TOSHIBA KK (TOKE )

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 2001155020	A	20010608	JP 99334597	A	19991125	200148 B

Priority Applications (No Type Date): JP 99334597 A 19991125

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
JP 2001155020	A	14	G06F-017/30	

Abstract (Basic): JP 2001155020 A

NOVELTY - Preclassified subjectwise and held over database are sought to be isolated those that have adequate **similarity** of contents compared to that of **search document** , from among host of **documents** . Frequency of **keywords** is determined in **documents** to be screened frequencies adjusted for size of **documents** . Graded **weightage** is assigned to **keywords** based on determined frequency and serves as basis for **document** selection.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(a) **Similar document search** procedure;

(b) Recording medium

USE - Databases holding vast number of documents need to be scanned to **locate** subject specific information available with select **document** .

ADVANTAGE - Leads to higher accuracy in the **location** of **documents** that have contents bearing adequate **similarity** to those obtaining with current **search document** .

DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of components of **similar document search** device. (Drawing includes non-English language text).

pp; 14 DwgNo 2/32

Title Terms: **DOCUMENT** ; SCREEN; BASED; **SIMILAR** ; CONTENT; INTRINSIC;  
**SEARCH** ; **DOCUMENT** ; DETERMINE; FREQUENCY; **KEYWORD** ; SCREEN; **DOCUMENT**

Derwent Class: T01

International Patent Class (Main): **G06F-017/30**

File Segment: EPI

19/5/25 (Item 7 from file: 350)  
DIALOG(R) File 350:Derwent WPIX  
(c) 2004 Thomson Derwent. All rts. reserv.

013251186 \*\*Image available\*\*  
WPI Acc No: 2000-423069/200036  
XRPX Acc No: N00-315724

**Characterizing term extraction method in computer, involves sorting  
extracted terms according to generated moduli and accepting terms with  
greatest moduli as characteristic keyword of documents content**  
Patent Assignee: JUSTSYSTEM PITTSBURGH RES CENT INC (JUST-N)  
Inventor: KANTROWITZ M  
Number of Countries: 090 Number of Patents: 002  
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
WO 200033215	A1	20000608	WO 99US25686	A	19991101	200036 B
AU 200019073	A	20000619	AU 200019073	A	19991101	200044

Priority Applications (No Type Date): US 98201569 A 19981130  
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
WO 200033215	A1	E	16	G06F-017/30	

Designated States (National): AE AL AM AT AU AZ BA BB BG BR BY CA CH CN  
CR CU CZ DE DK DM EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP  
KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE  
SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW

Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR  
IE IT KE LS LU MC MW NL OA PT SD SE SL SZ TZ UG ZW

AU 200019073 A G06F-017/30 Based on patent WO 200033215

Abstract (Basic): WO 200033215 A1

NOVELTY - Occurrences of each term extracted from **document** is counted to establish a frequency value for each term. The characters in each term is counted. The frequency value for each term or monotonic function is multiplied by character count or monotonic function to form modulus for each term. The terms are sorted according to the moduli and moduli is accepted as **characteristic keyword** of the **document**'s content.

USE - In computer, world wide web for term **weighting**, for information **retrieval** applications such as **document retrieval**, cross-language information **retrieval**, **keyword extraction**, **document routing**, classification, categorization, clustering, **document filtering**, **query expansion**, chapter, paragraph and sentence segmentation, spelling correction, term, **query** and **document similarity** metrics and text summarization.

ADVANTAGE - Size of indexes in the information **retrieval** algorithm is reduced. **Document** summarized is easy to implement and use and requires only less memory. The method is scalable because it does not rely on information outside the **document** and so does not consume more resources as the number of **documents** increases. So the method is highly suitable for distributed information **retrieval** applications.

DESCRIPTION OF DRAWING(S) - The figure shows the flow diagram explaining the computer program for implementing the characterizing terms extraction method.

pp; 16 DwgNo 1/1

Title Terms: TERM; EXTRACT; METHOD; COMPUTER; SORT; EXTRACT; TERM; ACCORD;  
GENERATE; MODULUS; ACCEPT; TERM; GREATER; MODULUS; **CHARACTERISTIC** ;  
**KEYWORD** ; **DOCUMENT** ; CONTENT

Derwent Class: T01  
International Patent Class (Main): **G06F-017/30**  
File Segment: EPI



Set	Items	Description
S1	9991	(INDEX OR THESAURUS OR KEY) ( ) (WORD? OR TERM? OR PHRASE?) OR KEYWORD? OR KEYTERM?
S2	1152186	SEARCH? OR SEEK? OR FIND? OR QUER? OR RETRIEV? OR LOCAT?
S3	1006393	AGENT? OR IA OR SPIDER? OR CRAWLER? OR WEBCRAWLER? OR BOT - OR (SOFTWARE?) ( ) (ROBOT?) OR SOFTBOT? OR BOTS
S4	283095	FILE? OR DOCUMENT? OR DATAFILE? OR ELECTRONIC()TEXT? OR ET-EXT? OR PAGE?
S5	1033005	RELEVAN? OR RANK? OR WEIGH? OR SCORE? OR POINTS
S6	2515527	SIMILAR? OR SAME? OR CONGRUENT? OR IDENTICAL? OR CHARACTER-ISTIC? OR FEATUR?
S7	21	S1 AND S2 AND S3 AND S4
S8	8	S7 AND (S5 OR S6)
S9	9	S7 AND IC=(G06F-015? OR G06F-007?)
S10	14	S8 OR S9
S11	14	IDPAT (sorted in duplicate/non-duplicate order)
S12	13	IDPAT (primary/non-duplicate records only)

? show files

File 347:JAPIO Oct 1976-2003/Sep(Updated 040105)

(c) 2004 JPO & JAPIO

File 350:Derwent WPIX 1963-2004/UD,UM &UP=200408

(c) 2004 Thomson Derwent

12/5/1 (Item 1 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2004 Thomson Derwent. All rts. reserv.

015770961 \*\*Image available\*\*  
WPI Acc No: 2003-833163/200377  
Related WPI Acc No: 2002-698395  
XRPX Acc No: N03-666117

**Web search engine e.g. for Google search engine ranks hyper text pages that are stored along with keyword in index database, by determining intrinsic and extrinsic ranks according to web page content and connectivity analysis**

Patent Assignee: CHUNG S (CHUN-I); DOD A (DODA-I); KIM B S (KIMB-I); KIM M (KIMM-I); YUN Y (YUNY-I)

Inventor: CHUNG S; DOD A; KIM B S; KIM M; YUN Y

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20030208482	A1	20031106	US 2001757435	A	20010110	200377 B
			US 2003454452	A	20030603	

Priority Applications (No Type Date): US 2001757435 A 20010110; US 2003454452 A 20030603

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
US 20030208482	A1	15	G06F-007/00	Div ex application US 2001757435

Abstract (Basic): US 20030208482 A1

NOVELTY - A **crawler** (12) fetches **pages** from web (13) and stores in a database (14). A URL management system (18) assigns identification number to the URL of each **page**. A indexer (26) parses **keyword** from the **pages** and stores the URL along with **keywords** in index database (28). A **ranking** unit (30) **ranks** the hypertext **pages** based on intrinsic or extrinsic **rank** provided to the **page** according to the content and connectivity analysis.

DETAILED DESCRIPTION - An INDEPENDENT CLAIM is also included for computer system.

USE - Web **search** engine such as Google, FAST, Altavista, Excite, Yahoo, HotBot, Infoseek and Northern light **search** engine.

ADVANTAGE - The most **relevant pages** of the **search** result is provided to the user, by assigning **ranks** to the hypertext **pages** for multi- **keyword query**.

DESCRIPTION OF DRAWING(S) - The figure shows the architecture of the **search** engine.

**crawler** (12)  
web (13)  
URL management system (18)  
indexer (26)  
indexed database (28)  
**ranker** (30)  
pp; 15 DwgNo 1/6

Title Terms: WEB; **SEARCH** ; ENGINE; **SEARCH** ; ENGINE; **RANK** ; HYPER; TEXT; **PAGE** ; STORAGE; **KEYWORD** ; INDEX; DATABASE; DETERMINE; INTRINSIC; EXTRINSIC; **RANK** ; ACCORD; WEB; **PAGE** ; CONTENT; CONNECT; ANALYSE

Derwent Class: T01

International Patent Class (Main): G06F-007/00

File Segment: EPI

12/5/2 (Item 2 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2004 Thomson Derwent. All rts. reserv.

015770959 \*\*Image available\*\*  
WPI Acc No: 2003-833161/200377  
XRPX Acc No: N03-666115

Document keywords **linking method involves applying hot key on selected keyword , to trigger access agent that retrieves uniform resource locator which is used to invoke web browser**

Patent Assignee: PHAM P M (PHAM-I)

Inventor: PHAM P M

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20030208472	A1	20031106	US 2000547117	A	20000411	200377 B
			US 2003427082	A	20030123	

Priority Applications (No Type Date): US 2003427082 A 20030123; US 2000547117 A 20000411

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 20030208472	A1	29	G06F-015/16	CIP of application	US 2000547117

Abstract (Basic): US 20030208472 A1

NOVELTY - The hot keys mapped to several uniform resource locators (URLs), are applied to a selected **keyword** that is displayed on a **document** viewer. The keys trigger an access **agent** to **retrieve** URL associated with the hot key and the user selected **keyword**. The **agent** replaces the marker in the URL with the selected **keyword**, and invokes web browser on the user system, by passing URL as command argument.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

- (1) method for hyperlinking hypertext in web **page**; and
- (2) **document keywords** linking system.

USE - For transparently linking **keywords** of **document** displayed in **document** viewer of user computer, to web sites offering **keyword** based information look-up services.

ADVANTAGE - The user's interaction in **retrieving** desired information on world wide web associated with **keyword** displayed in viewer, is optimized effectively. Helps Internet user's to **search** information from several information sources easily and lost link problem are eliminated effectively.

DESCRIPTION OF DRAWING(S) - The figure shows a schematic view of the end user system connected to web site through network.

Internet (20)  
web site (22)  
end user computer (30)  
pp; 29 DwgNo 1/14

Title Terms: **DOCUMENT ; KEYWORD ; LINK; METHOD; APPLY; HOT; KEY; SELECT; KEYWORD ; TRIGGER; ACCESS; AGENT ; RETRIEVAL ; UNIFORM; RESOURCE; LOCATE ; INVOKE; WEB**

Derwent Class: T01

International Patent Class (Main): **G06F-015/16**

International Patent Class (Additional): **G06F-007/00 ; G06F-017/00**

File Segment: EPI

12/5/4 (Item 4 from file: 350)  
DIALOG(R) File 350:Derwent WPIX  
(c) 2004 Thomson Derwent. All rts. reserv.

015272451 \*\*Image available\*\*  
WPI Acc No: 2003-333380/200331  
XRPX Acc No: N03-267186

**Database query and information delivery method for obtaining and displaying information to user that is of interest to individual associated with user**

Patent Assignee: LUCAS S (LUCA-I); MOORE E (MOOR-I); CLIENTELLIGENCE INC (CLIE-N)

Inventor: LUCAS S; MOORE E

Number of Countries: 101 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
WO 200327907	A1	20030403	WO 2002US31092	A	20020925	200331 B
US 20030065649	A1	20030403	US 2001966355	A	20010928	200340

Priority Applications (No Type Date): US 2001966355 A 20010928

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
WO 200327907	A1	E	44	G06F-017/30	

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG UZ VC VN YU ZA ZM ZW

Designated States (Regional): AT BE BG CH CY CZ DE DK EA EE ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SK SL SZ TR TZ UG ZM ZW  
US 20030065649 A1 G06F-007/00

Abstract (Basic): WO 200327907 A1

NOVELTY - A user is enabled to identify interested individuals and electronically mail (62) **relevant** objects of interest to selected individuals or to add objects to an action list.

DETAILED DESCRIPTION - The method involves importing or creating individual profiles with associated **key - phrases**, and **querying** all sources of information on a network (48) based on user-based or individual-based **key - phrases** with **relevant** objects e.g. **documents**, **retrieved** based on **key - phrase** occurrence and based on association to individuals who have matching **key - phrases**. This allows a user to easily identify interested individuals and electronically mail (62) **relevant** objects of interest to selected individuals or to add objects to an action list.

INDEPENDENT CLAIMS are included for; a system for providing information that is of interest to a group of individuals associated with the user; a method for identifying in a group individuals at least one that has an interest in information that a user possesses.

USE - Providing a user with information that is of interest to a group of individuals associated with the user; a system for providing a user with information that is of interest to a group of individuals associated with the user. Exchange of timely and **relevant** communication e.g. between; brokers, **agents**, sales professionals, stock brokers, financial advisers, real estate **agents**, travel **agents**, insurance **agents** etc.

ADVANTAGE - Provides user e.g. broker or **agent** with solution to improving and enabling communication with their clients using algorithms which cross-reference the interests of clients with any information pool and present a list of clients who are interested in

the **relevant** information.

DESCRIPTION OF DRAWING(S) - The drawing shows a schematic diagram illustrating the general method of the invention.

User broker (40)

User interface (42)

Application database (44)

Application **search agent** (46)

Information provider (52,54)

**Search** engines (56)

Individual client (58)

Web browser (60)

Email (62)

pp; 44 DwgNo 2/7

Title Terms: DATABASE; **QUERY** ; INFORMATION; DELIVER; METHOD; OBTAIN;  
DISPLAY; INFORMATION; USER; INTEREST; INDIVIDUAL; ASSOCIATE; USER

Derwent Class: T01

International Patent Class (Main): **G06F-007/00** ; G06F-017/30

File Segment: EPI

12/5/8 (Item 8 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2004 Thomson Derwent. All rts. reserv.

013980776 \*\*Image available\*\*  
WPI Acc No: 2001-464990/200150  
Related WPI Acc No: 1998-467772; 2001-390074  
XRPX Acc No: N01-344924

Information access/ retrieval method for hand-held device by parsing  
input text expression to identify keyword

Patent Assignee: ACTIONEER INC (ACTI-N)  
Inventor: MARTIN F  
Number of Countries: 094 Number of Patents: 003  
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
WO 200142906	A1	20010614	WO 2000US33229	A	20001207	200150 B
AU 200120700	A	20010618	AU 200120700	A	20001207	200161
EP 1257905	A1	20021120	EP 2000984020	A	20001207	200301
			WO 2000US33229	A	20001207	

Priority Applications (No Type Date): US 99169539 P 19991207

Patent Details:

Patent No Kind Lan Pg Main IPC Filing Notes

WO 200142906 A1 E 31 G06F-007/10

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA  
CH CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP  
KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT  
RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG UZ VN YU ZA ZW

Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR  
IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW

AU 200120700 A G06F-007/10 Based on patent WO 200142906

EP 1257905 A1 E G06F-007/10 Based on patent WO 200142906

Designated States (Regional): AL AT BE CH CY DE DK ES FI FR GB GR IE IT  
LI LT LU LV MC MK NL PT RO SE SI TR

Abstract (Basic): WO 200142906 A1

NOVELTY - The input text expression is parsed to identify **keyword**, a connector that includes grammatical rules is determined based on the **keyword** and a template in the connector is filled based on the input text expression. An **agent** identified by the connector is then launched and the response received as streaming HTML data.

DETAILED DESCRIPTION - An INDEPENDENT claim is also included for USE - For hand-held device.

ADVANTAGE - It allows the hand-held device to access/ **retrieve** information from the Internet without requiring processing HTML **pages** remotes using a proxy server that requires maintenance.

DESCRIPTION OF DRAWING(S) - The figure shows flow chart of information access/ **retrieval** method.

pp; 31 DwgNo 4A/7

Title Terms: INFORMATION; ACCESS; **RETRIEVAL** ; METHOD; HAND; HELD; DEVICE;  
PARSE; INPUT; TEXT; EXPRESS; IDENTIFY; **KEYWORD**

Derwent Class: T01

International Patent Class (Main): G06F-007/10

International Patent Class (Additional): G06F-015/16 ; G06F-017/30

File Segment: EPI

12/5/9 (Item 9 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2004 Thomson Derwent. All rts. reserv.

013825395 \*\*Image available\*\*  
WPI Acc No: 2001-309607/200133  
XRPX Acc No: N01-221614

**Apparatus for web information extraction services using a client module with means to search web sites corresponding to target uniform resource locators**

Patent Assignee: UNIV INFORMATION & COMMUNICATIONS (UYIN-N); INFORMATION & COMMUNICATIONS FOUND (INFO-N)

Inventor: HYUN S J; KIM G B

Number of Countries: 026 Number of Patents: 003

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
EP 1069515	A1	20010117	EP 2000115316	A	20000714	200133 B
KR 2001009983	A	20010205	KR 9928638	A	19990715	200152
KR 359233	B	20021101	KR 9928638	A	19990715	200329

Priority Applications (No Type Date): KR 9928638 A 19990715

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
EP 1069515	A1	E 12	G06F-017/30	

Designated States (Regional): AL AT BE CH CY DE DK ES FI FR GB GR IE IT  
LI LT LU LV MC MK NL PT RO SE SI

KR 2001009983	A	G06F-017/30
---------------	---	-------------

KR 359233	B	G06F-017/30	Previous Publ. patent KR 2001009983
-----------	---	-------------	-------------------------------------

Abstract (Basic): EP 1069515 A1

NOVELTY - Web **pages** in hyper space (10) may carry information **relevant** to a **keyword** given by a web user and a web information extractor provides the server with an extracted result (20) in the form of a single user **file** obtained according to three main functions, I.e. hyper-link traversal for **finding** desired information **pages**, **searching** and collecting them into a user **file**. The functions are performed according to a target uniform resource **locator** provided from a **search** engine.

DETAILED DESCRIPTION - AN INDEPENDENT CLAIM is included for a method for web information extraction in an intelligent **agent** system.

USE - Web information extraction.

ADVANTAGE - Automatic hyper-link space **searching** and quick content collection.

DESCRIPTION OF DRAWING(S) - The drawing is a schematic diagram of the overall concept of a web information extractor

Hyper space (10)

Extracted result (20)

pp; 12 DwgNo 1/4

Title Terms: APPARATUS; WEB; INFORMATION; EXTRACT; SERVICE; CLIENT; MODULE;  
**SEARCH** ; WEB; SITE; CORRESPOND; TARGET; UNIFORM; RESOURCE; **LOCATE**

Derwent Class: T01

International Patent Class (Main): G06F-017/30

File Segment: EPI

12/5/11 (Item 11 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2004 Thomson Derwent. All rts. reserv.

012737347 \*\*Image available\*\*  
WPI Acc No: 1999-543464/199946  
XRPX Acc No: N99-403064

**Web search engine using user background information**

Patent Assignee: MITEL CORP (MTLC )  
Inventor: WEISS M  
Number of Countries: 003 Number of Patents: 004  
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
GB 2335761	A	19990929	GB 986392	A	19980325	199946 B
DE 19913509	A1	19990930	DE 1013509	A	19990325	199946
CA 2265292	A1	19990925	CA 2265292	A	19990315	200008
GB 2335761	B	20030514	GB 986392	A	19980325	200333

Priority Applications (No Type Date): GB 986392 A 19980325

**Patent Details:**

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
GB 2335761	A		24	G06F-017/30	
DE 19913509	A1		13	G06F-017/30	
CA 2265292	A1 E			G06F-013/14	
GB 2335761	B			G06F-017/30	

Abstract (Basic): GB 2335761 A

NOVELTY - The **search** system uses a web **agent** (126) to manage a portfolio of web **page** profiles for user's (118, 120) stored in the server (112) memory (116). The user profile contains background information on the user with information about the user's interest in particular **features** and web **pages**.

DETAILED DESCRIPTION - The user profile can be generated by either manual entry or a learning process based on which sites the user visits and length of time spent at that site.

When carrying out a **search** for **keywords** the user profile is first used to select **pages** of interest to the user then these **pages** are further filtered by removing those without the **keyword** before the results are displayed.

USE - For use as a **search** engine for the world wide web.

ADVANTAGE - The addition of the user profile to the **search** engine gives a more specific list of web sites that the user may be interested in than a **keyword search** alone would provide, lowering the amount of sites returned by the **search** engine.

DESCRIPTION OF DRAWING(S) - Block diagram of a network system implementing the user profile based **search**.

Web server (112)  
Web server memory element (116)  
Web browser (user) (118, 120)  
Web **agent** for **search** engine (126)  
pp; 24 DwgNo 1/4

Title Terms: WEB; **SEARCH** ; ENGINE; USER; BACKGROUND; INFORMATION  
Derwent Class: T01  
International Patent Class (Main): G06F-013/14; G06F-017/30  
File Segment: EPI



12/5/13 (Item 13 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

010865894 \*\*Image available\*\*

WPI Acc No: 1996-362845/199636

Related WPI Acc No: 1996-322055

XRPX Acc No: N96-305825

Information access system for accessing distributed information - uses distributed system of intelligent software agents to perform information tasks on behalf of user community, e.g store, retrieve, summarise and inform other agents about information found

Patent Assignee: BRITISH TELECOM PLC (BRTE )

Inventor: DAVIES N J; WEEKS R

Number of Countries: 071 Number of Patents: 018

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
WO 9623265	A1	19960801	WO 96GB132	A	19960123	199636	B
AU 9644549	A	19960814	AU 9644549	A	19960123	199650	
			WO 96GB132	A	19960123		
FI 9703080	A	19970722	WO 96GB132	A	19960123	199743	
			FI 973080	A	19970722		
NO 9703372	A	19970922	WO 96GB132	A	19960123	199749	
			NO 973372	A	19970722		
EP 807291	A1	19971119	EP 96900645	A	19960123	199751	
			WO 96GB132	A	19960123		
BR 9606931	A	19971111	BR 966931	A	19960123	199801	
			WO 96GB132	A	19960123		
MX 9705582	A1	19971101	MX 975582	A	19970723	199902	
JP 10513587	W	19981222	JP 96522713	A	19960123	199910	
			WO 96GB132	A	19960123		
NZ 298861	A	19990128	NZ 298861	A	19960123	199910	
			WO 96GB132	A	19960123		
KR 98701598	A	19980515	WO 96GB132	A	19960123	199918	
			KR 97704990	A	19970723		
AU 707050	B	19990701	AU 9644549	A	19960123	199937	
US 5931907	A	19990803	WO 96GB132	A	19960123	199937	
			US 97875091	A	19970722		
EP 953920	A2	19991103	EP 96900645	A	19960123	199951	
			EP 99113304	A	19960123		
EP 807291	B1	20000105	EP 96900645	A	19960123	200006	
			WO 96GB132	A	19960123		
			EP 99113304	A	19960123		
DE 69606021	E	20000210	DE 606021	A	19960123	200015	
			EP 96900645	A	19960123		
			WO 96GB132	A	19960123		
US 6289337	B1	20010911	WO 96GB132	A	19960123	200154	
			US 97875091	A	19970722		
			US 99351633	A	19990712		
CN 1169195	A	19971231	CN 96191566	A	19960123	200168	
CA 2210581	C	20020326	CA 2210581	A	19960123	200230	
			WO 96GB132	A	19960123		

Priority Applications (No Type Date): EP 95300420 A 19950123; WO 95GB3017 A 19951221

Cited Patents: 05Jnl.Ref; EP 361464

Patent Details:

Patent No Kind Lan Pg Main IPC Filing Notes

WO 9623265 A1 E 25 G06F-017/30

Designated States (National): AL AM AT AU AZ BB BG BR BY CA CH CN CZ DE DK EE ES FI GB GE HU IS JP KE KG KP KR KZ LK LR LS LT LU LV MD MG MK MN

MW MX NO NZ PL PT RO RU SD SE SG SI SK TJ TM TR TT UA UG US UZ VN  
 Designated States (Regional): AT BE CH DE DK EA ES FR GB GR IE IT KE LS  
 LU MC MW NL OA PT SD SE SZ UG

AU 9644549	A	G06F-017/30	Based on patent WO 9623265
FI 9703080	A	G06F-000/00	
NO 9703372	A	G06F-017/30	
EP 807291	A1 E	G06F-017/30	Based on patent WO 9623265
Designated States (Regional): BE CH DE DK ES FR GB IT LI NL PT SE			
BR 9606931	A	G06F-017/30	Based on patent WO 9623265
MX 9705582	A1	G06F-017/30	
JP 10513587	W	32 G06F-017/30	Based on patent WO 9623265
NZ 298861	A	G06F-017/30	Based on patent WO 9623265
KR 98701598	A	G06F-017/30	Based on patent WO 9623265
AU 707050	B	G06F-017/30	Previous Publ. patent AU 9644549
US 5931907	A	G06F-017/30	Based on patent WO 9623265
EP 953920	A2 E	G06F-017/30	Div ex application EP 96900645
			Div ex patent EP 807291
Designated States (Regional): AT BE CH DE DK ES FR GB GR IE IT LI LT LU			
EP 807291	B1 E	G06F-017/30	Related to application EP 99113304
			Related to patent EP 953920
			Based on patent WO 9623265
Designated States (Regional): BE CH DE DK ES FR GB IT LI NL PT SE			
DE 69606021	E	G06F-017/30	Based on patent EP 807291
			Based on patent WO 9623265
US 6289337	B1	G06F-007/00	Div ex application WO 96GB132
			Div ex application US 97875091
			Div ex patent US 5931907
CN 1169195	A	G06F-017/30	
CA 2210581	C E	G06F-017/30	Based on patent WO 9623265

Abstract (Basic): WO 9623265 A

The system for accessing information stored in a distributed information database uses a community of intelligent software **agents** (105), each of which can be built as an extension of a known viewer (400) for a distribution information system e.g Internet World Wide Web. The **agent** is integrated with the viewer (400) and can extract **pages** by means of the viewer (400) for storage in an intelligent **page** store.

The text from the information system is abstracted and stored with additional information, selected by the user. The **agent** based system uses **keyword** sets to **locate** information of interest to the user, together with user profiles, such that **pages** being stored by one user can be notified to another user whose profile indicates potential interest.

USE - **Locating** information on e.g Internet, HyperText **documents** **located** on user's internal systems etc.

Dwg.1/9

Title Terms: INFORMATION; ACCESS; SYSTEM; ACCESS; DISTRIBUTE; INFORMATION; DISTRIBUTE; SYSTEM; INTELLIGENCE; SOFTWARE; **AGENT** ; PERFORMANCE; INFORMATION; TASK; USER; COMMUNAL; STORAGE; **RETRIEVAL** ; SUMMARY; INFORMATION; **AGENT** ; INFORMATION; FOUND

Derwent Class: T01

International Patent Class (Main): G06F-000/00; **G06F-007/00** ; G06F-017/30

International Patent Class (Additional): **G06F-015/00**

File Segment: EPI

Set	Items	Description
S1	23998	(INDEX OR THESAURUS OR KEY) ( ) (WORD? OR TERM? OR PHRASE?) OR KEYWORD? OR KEYTERM?
S2	3742197	SEARCH? OR SEEK? OR FIND? OR QUER? OR RETRIEV? OR LOCAT?
S3	1389449	AGENT? OR IA OR SPIDER? OR CRAWLER? OR WEBCRAWLER? OR BOT - OR (SOFTWARE?) ( ) (ROBOT?) OR SOFTBOT? OR BOTS
S4	1139670	FILE? OR DOCUMENT? OR DATAFILE? OR ELECTRONIC ( ) TEXT? OR ET- EXT? OR PAGE?
S5	2759322	RELEVAN? OR RANK? OR WEIGH? OR SCORE? OR POINTS
S6	7724291	SIMILAR? OR SAME? OR CONGRUENT? OR IDENTICAL? OR CHARACTER- ISTIC? OR FEATUR?
S7	38	S1 AND S2 AND S3 AND S4 AND S5 AND S6
S8	902	S1 AND S2 AND S4 (4N) (S5 OR S6)
S9	575	S1 AND S2 AND S4 (3N) S5
S10	305	S1 AND S2 AND S4 (3N) S6
S11	64	S9 AND S10
S12	99	S7 OR S11
S13	73	RD (unique items)
S14	55	S13 NOT PY>2001
S15	54	S14 NOT PD=20010116:20030116
S16	54	S15 NOT PD=20030116:20040211

? show files

File 8: Ei Compendex(R) 1970-2004/Jan W3  
(c) 2004 Elsevier Eng. Info. Inc.

File 35: Dissertation Abs Online 1861-2004/Jan  
(c) 2004 ProQuest Info&Learning

File 202: Info. Sci. & Tech. Abs. 1966-2004/Jan 20  
(c) 2004 EBSCO Publishing

File 65: Inside Conferences 1993-2004/Feb W1  
(c) 2004 BLDSC all rts. reserv.

File 2: INSPEC 1969-2004/Jan W4  
(c) 2004 Institution of Electrical Engineers

File 94: JICST-EPlus 198  
(c) 2004 Japan S

File 111: TGG Natl. Newspa. Jan 29  
(c) 2004 The Ga

File 233: Internet & Pers 3/Sep  
(c) 2003 EBSCO

File 144: Pascal 1973-200  
(c) 2004 INIST/

File 34: SciSearch(R) Ci n W4  
(c) 2004 Inst f

File 99: Wilson Appl. Sci & Techn Abs 1983-2004/Dec  
(c) 2004 The HW Wilson Co.

*Bibliography*

*Databases*

16/5/1 (Item 1 from file: 8)  
DIALOG(R)File 8:Ei Compendex(R)  
(c) 2004 Elsevier Eng. Info. Inc. All rts. reserv.

05866180 E.I. No: EIP01316599389

**Title: Features: Real-time adaptive feature and document learning for web search**

Author: Chen, Z.; Meng, X.; Fowler, R.H.; Zhu, B.

Corporate Source: Dept. of Computer Science Univ. of Texas-Pan American, Edinburg, TX 78539-2999, United States

Source: Journal of the American Society for Information Science and Technology v 52 n 8 June 2001. p 655-665

Publication Year: 2001

CODEN: AISJB6 ISSN: 1532-2882

Language: English

Document Type: JA; (Journal Article) Treatment: A; (Applications); T; (Theoretical)

Journal Announcement: 0108W1

Abstract: In this article we report our research on building Features-an intelligent web **search** engine that is able to perform real-time adaptive feature (i.e., **keyword**) and document learning. Not only does Features learn from the user's **document relevance** feedback, but it also automatically extracts and suggests indexing **keywords** relevant to a search query and learns from the user's **keyword** relevance feedback so that it is able to speed up its **search** process and to enhance its **search** performance. We design two efficient and mutual-benefiting learning algorithms that work concurrently, one for feature learning and the other for **document** learning. **Features** employs these algorithms together with an internal index database and a real-time meta-**searcher** to perform adaptive real-time learning to **find** desired **documents** with as little **relevance** feedback from the user as possible. The architecture and performance of Features are also discussed. 29 Refs.

Descriptors: **Search** engines; World Wide Web; Learning systems; Adaptive systems; Real time systems; Information **retrieval** systems; Automatic indexing; Learning algorithms

Identifiers: Intelligent web **search** engine; **Feature** learning; **Document** learning; **Relevance** feedback

Classification Codes:

723.5 (Computer Applications); 723.4 (Artificial Intelligence); 722.4 (Digital Computers & Systems); 903.3 (Information Retrieval & Use); 903.1 (Information Sources & Analysis)

723 (Computer Software, Data Handling & Applications); 722 (Computer Hardware); 903 (Information Science)

72 (COMPUTERS & DATA PROCESSING); 90 (ENGINEERING, GENERAL)

16/5/4 (Item 4 from file: 8)

DIALOG(R) File 8: Ei Compendex(R)

(c) 2004 Elsevier Eng. Info. Inc. All rts. reserv.

05051028 E.I. No: EIP98074264818

**Title:** WebMate: A personal agent for browsing and searching

**Author:** Chen, Liren; Sycara, Katia

**Corporate Source:** Carnegie Mellon Univ, Pittsburgh, PA, USA

**Conference Title:** Proceedings of the 1998 2nd International Conference on Autonomous Agents

**Conference Location:** Minneapolis, MN, USA **Conference Date:** 19980509-19980513

**Sponsor:** ACM

**E.I. Conference No.:** 48571

**Source:** Proceedings of the Interantional Conference on Autonomous Agents 1998. ACM, New York, NY, USA. p 132-139

**Publication Year:** 1998

**CODEN:** 002624

**Language:** English

**Document Type:** CA; (Conference Article) **Treatment:** T; (Theoretical)

**Journal Announcement:** 9808W4

**Abstract:** The World-Wide Web is developing very fast. Currently, **finding** useful information on the Web is a time consuming process. In this paper, we present WebMate, an **agent** that helps users to effectively browse and **search** the Web. WebMate extends the state of the art in Web-based information **retrieval** in many ways. First, it uses multiple TF-IDF vectors to keep track of user interests in different domains. These domains are automatically learned by WebMate. Second, WebMate uses the Trigger Pair Model to automatically extract **keywords** for refining **document search**. Third, during **search**, the user can provide multiple **pages** as **similarity / relevance** guidance for the **search**. The system extracts and combines **relevant keywords** from these **relevant pages** and uses them for **keyword** refinement. Using these techniques, WebMate provides effective browsing and **searching** help and also compiles and sends to users personal newspaper by automatically spiding news sources. We have experimentally evaluated the performance of the system. (Author abstract) 19 Refs.

**Descriptors:** Artificial intelligence; Computer aided software engineering ; **Query** languages; Wide area networks; Online **searching** ; Information **retrieval** systems

**Identifiers:** World wide web (WWW)

**Classification Codes:**

723.4 (Artificial Intelligence); 723.1 (Computer Programming); 723.5 (Computer Applications); 723.3 (Database Systems); 722.4 (Digital Computers & Systems)

723 (Computer Software); 722 (Computer Hardware)

72 (COMPUTERS & DATA PROCESSING)

16/5/5 (Item 5 from file: 8)

DIALOG(R) File 8: Ei Compendex(R)

(c) 2004 Elsevier Eng. Info. Inc. All rts. reserv.

04755885 E.I. No: EIP97073739834

**Title: Musag: An agent that learns what you mean**

Author: Goldman, Claudia V.; Langer, Amir; Rosenschein, Jeffrey S.

Corporate Source: Hebrew Univ, Jerusalem, Isr

Source: Applied Artificial Intelligence v 11 n 5 Jul-Aug 1997. p 413-435

Publication Year: 1997

CODEN: AAINEH ISSN: 0883-9514

Language: English

Document Type: JA; (Journal Article) Treatment: G; (General Review)

Journal Announcement: 9709W2

**Abstract:** A system that carries out highly effective **searches** over collections of textual information is presented. The system is made up of two major parts. The first part consists of an **agent**, Musag, that learns to relate concepts that are semantically '**similar**' to one another. The second part consists of another **agent**, Sag, which is responsible for **retrieving documents**, given a set of **keywords** with relative **weights**. The **agents**' system architecture, along with the nature of their interactions, the learning and **search** algorithms, the notion of 'cost of learning' and how it influences the learning process and the quality of the dictionary at any given time are described. 8 Refs.

**Descriptors:** Learning systems; **Query** languages; Information **retrieval** systems; User interfaces; Computer software; Computer architecture; Learning algorithms; Linguistics

**Identifiers:** Syntactic method; Software package Musag

**Classification Codes:**

723.4.1 (Expert Systems)

723.4 (Artificial Intelligence); 723.3 (Database Systems); 903.3

(Information Retrieval & Use); 722.2 (Computer Peripheral Equipment)

723 (Computer Software); 903 (Information Science); 722 (Computer Hardware)

72 (COMPUTERS & DATA PROCESSING); 90 (GENERAL ENGINEERING)

16/5/9 (Item 9 from file: 8)

DIALOG(R)File 8: Ei Compendex(R)

(c) 2004 Elsevier Eng. Info. Inc. All rts. reserv.

02358459 E.I. Monthly No: EIM8712-085931

**Title:** DOCUMENT - DOCUMENT SIMILARITY MEASURE BASED ON CITED TITLES  
AND PROBABILITY THEORY, AND ITS APPLICATION TO RELEVANCE FEEDBACK  
RETRIEVAL .

**Author:** Kwok, K. L.

**Corporate Source:** Queens Coll, Flushing, NY, USA

**Conference Title:** Research and Development in Information Retrieval,  
Proceedings of the Third Joint BCS and ACM Symposium.

**Conference Location:** Cambridge, Engl **Conference Date:** 19840702

**Sponsor:** British Computer Soc, London, Engl; ACM, New York, NY, USA

**E.I. Conference No.:** 10440

**Source:** Publ by Cambridge Univ Press (British Computer Society Workshop  
Series), Cambridge, Engl and New York, NY, USA p 221-231

**Publication Year:** 1984

**ISBN:** 0-521-26865-6

**Language:** English

**Document Type:** PA; (Conference Paper)

**Journal Announcement:** 8712

**Abstract:** The use of cited title terms of a scientific document for automatic indexing is explored. It offers a means of **index term** selection as well as term relevance weighting, based on author-provided relevance information and Bayes Theorem as in probabilistic **retrieval** . The latter quantitative consideration leads to a new measure of **document - document similarity** measure which is shown to have importance both for initial **search** and in relevance feedback **retrieval** , by offering a choice of iterative strategies. Extension of the concept of cited title terms to citing title terms shows that these two approaches are compatible with the current two competing models of probability of **relevance** for **document retrieval** (Robertson et al. 1982), if a document can also be regarded as a **query** . Their term usage may therefore provide the necessary statistics for parameter estimation to test both theories. (Author abstract) 17 refs.

**Descriptors:** INFORMATION SCIENCE--\*Indexing; INFORMATION **RETRIEVAL**  
SYSTEMS--Evaluation; DATABASE SYSTEMS-- **Query** Languages

**Identifiers:** **DOCUMENT - DOCUMENT SIMILARITY MEASURE; RELEVANCE**  
**FEEDBACK RETRIEVAL ; CITED TITLE TERMS; TERM RELEVANCE WEIGHTING; BAYES**  
**THEOREM; CITATION INDEXING**

**Classification Codes:**

903 (Information Science); 723 (Computer Software)

90 (GENERAL ENGINEERING); 72 (COMPUTERS & DATA PROCESSING)

16/5/18 (Item 6 from file: 202)

DIALOG(R) File 202:Info. Sci. & Tech. Abs.  
(c) 2004 EBSCO Publishing. All rts. reserv.

0100163

**Mathematical theories of relevance with respect to the problems of indexing.**

Book Title: An Algorithm For Document Characterization. Rept. No. R-2. Grant No. Nsf Gn177. 1965 March 12. Center For The Information Sciences, Lehigh University, Bethlehem, Pa. 59 P. Cfsti, Pb-167-444, Hc \$3.00, Mf \$0.75.

Author(s): Hillman, Donald J  
Corporate Source: Lehigh Univ  
Publication Date: 1965  
Language: English  
Document Type: Book Chapter  
Record Type: Abstract  
Journal Announcement: 0100

A distinction is made between document **retrieval** systems and 'fact **retrieval** ' systems. It is stipulated that for the former the **index terms** should be the names of the topics dealt with by the documents in the system collection. Such **index terms** are called ' **document characteristics** .' a **document** is then regarded as a complex assertion, and the problem of discovering its characteristics is defined to be that of isolating the referring expressions in the components of the complex assertion. It is shown that the type of reference discernable in simple sentences is preserved when such sentences are transformationally combined to produce complex sentences. Two methods of sentence reduction are examined for this purpose, vis., the derivation of microsentences and a kernelization program. Kernels are inefficient for document characterization purposes. Hence, an algorithm is constructed which operates on kernels to form certain micro-sentences called "assertive components". This algorithm together with a method for weighting the referring expressions of assertive components provide the means for assigning **characteristics** to any given **document** . The **characteristics** accurately denote the topics about which assertions have been made in the **document** , and the **weighting** of the characteristics supplies a means for assessing how much of the document's content is taken up with a discussion of those topics.

Classification Codes and Description: 4.07 (Classification, Indexing, and Thesauri)

Main Heading: Information Recognition and Description



IEE

16/5/22 (Item 4 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2004 Institution of Electrical Engineers. All rts. reserv.

6681817 INSPEC Abstract Number: C2000-10-7250R-001

**Title: Information gathering based on user's interest by multiple Web robots**

Author(s): Nagino, N.; Yamada, S.

Author Affiliation: CISS, Tokyo Inst. of Technol., Yokohama, Japan

Journal: Transactions of the Institute of Electronics, Information and Communication Engineers D-I vol.J83D-I, no.7 p.780-8

Publisher: Inst. Electron. Inf. & Commun. Eng,

Publication Date: July 2000 Country of Publication: Japan

CODEN: DTRDES ISSN: 0915-1915

SICI: 0915-1915(200007)J83DI:7L.780:IGBU;1-F

Material Identity Number: M972-2000-007

Language: Japanese Document Type: Journal Paper (JP)

Treatment: Practical (P)

**Abstract:** The authors propose a PWM (Personal Web Map) system which gathers information through the WWW depending on a user's interests. The PWM is a database of **relevant Web pages** constructed under his/her interactive control. For a user's easy understanding of the gathering process, the system controls Web robots to keep PWM as uniform as possible on **keywords**. The gathering process is indicated using a 2D map generated by SOM (self-organizing map), and a user gives the system feedback through it. Finally, we conducted various experiments, and proved that a PWM system was promising for information gathering in the WWW. (13 Refs)

Subfile: C

**Descriptors:** information resources; information **retrieval** ; online front-ends; self-organising **feature** maps; software **agents**

**Identifiers:** information gathering; user interest; multiple Web robots; PWM; Personal Web Map; WWW; **relevant** Web **pages** ; interactive control; **keywords** ; 2D map; SOM; self-organizing map; system feedback

**Class Codes:** C7250R (Information retrieval techniques); C7210N (Information networks); C7250N (Search engines); C6170 (Expert systems and other AI software and techniques); C5290 (Neural computing techniques)

Copyright 2000, IEE

16/5/27 (Item 9 from file: 2)

DIALOG(R)File 2:INSPEC

(c) 2004 Institution of Electrical Engineers. All rts. reserv.

6461583 INSPEC Abstract Number: C2000-02-7250R-045

**Title: Genetic programming for information retrieval**

Author(s): McKechnie, J.

Author Affiliation: Constr. Inf. Res. Centre, Newcastle upon Tyne Univ., UK

Conference Title: IEE Two-day Seminar. Searching for Information: Artificial Intelligence and Information Retrieval Approaches (Ref. No.1999/199) p.21/1-3

Publisher: IEE, London, UK

Publication Date: 1999 Country of Publication: UK 118 pp.

Material Identity Number: XX-1999-03515

Conference Title: IEE Two-day Seminar. Searching for Information: Artificial Intelligence and Information Retrieval Approaches

Conference Sponsor: IEE

Conference Date: 11-12 Nov. 1999 Conference Location: Glasgow, UK

Language: English Document Type: Conference Paper (PA)

Treatment: Practical (P)

Abstract: This paper discusses a new project that investigates the application of genetic programming to the generation of **index term0 weighting** schemes and **document - query similarity** measures for information **retrieval**. Despite the large number of existing schemes and measures that have been proposed, it is argued that varied and changing **retrieval** contexts justify the **search** for new methods. Preliminary work on the cystic fibrosis reference collection is described. (9 Refs)

Subfile: C

Descriptors: evolutionary computation; information **retrieval**; medical information systems; vocabulary

Identifiers: genetic programming; information **retrieval**; **index term** weighting schemes; **document - query similarity** measures; cystic fibrosis reference collection

Class Codes: C7250R (Information retrieval techniques); C7240 (Information analysis and indexing); C7140 (Medical administration)

Copyright 2000, IEE

16/5/30 (Item 12 from file: 2)

DIALOG(R) File 2:INSPEC

(c) 2004 Institution of Electrical Engineers. All rts. reserv.

6181620 INSPEC Abstract Number: C1999-04-7240-003

**Title: Statistical identification of domain-specific keyterms for text summarisation**

Author(s): Yuwono, B.; Adriani, M.

Author Affiliation: Graham Technol. plc, Glasgow, UK

Conference Title: Research and Advanced Technology for Digital Libraries.

Second European Conference, ECDL'98. Proceedings p.637-8

Editor(s): Nikolaou, C.; Stephanidis, C.

Publisher: Springer-Verlag, Berlin, Germany

Publication Date: 1998 Country of Publication: Germany xv+908 pp.

ISBN: 3 540 65101 2 Material Identity Number: XX-1998-02781

Conference Title: Reserch and Advanced Technology for Digital Libraries.

Second European Conference, ECDL'98. Proceedings

Conference Date: 21-23 Sept. 1998 Conference Location: Heraklion, Greece

Language: English Document Type: Conference Paper (PA)

Treatment: Practical (P)

**Abstract:** In order to be useful a text summarisation technique must be domain dependent, in that the resulting summary must cover the important aspects and concepts specific to the subject matter's domain. The main problem with a typical domain-dependent text summarisation technique is the cost of acquiring and hand-coding the required domain-specific knowledge into the system. To solve this problem, we propose a solution which uses automatically **retrieved** sample documents as the source of the domain-specific knowledge, and extracts the knowledge in the form of **key terms**. These **key terms** represent the key aspects and concepts (terminology) **relevant** to the input **document**. The sample documents are **retrieved** from a collection, called the base collection, containing documents of various topics, based on their **similarity** with the input **document**. The input document is then summarised by extracting a number of sentences containing the **key terms**. Our text summarisation technique is based on the statistical distribution of words among documents in the base collection, within individual documents, and among sentences in the input document. (0 Refs)

Subfile: C

Descriptors: abstracting; knowledge based systems; statistical analysis; vocabulary

Identifiers: statistical identification; domain-specific **key terms**; text summarisation; domain-dependent text summarisation; domain-specific knowledge; sample documents; **key terms**; terminology; base collection; statistical distribution

Class Codes: C7240 (Information analysis and indexing); C6170K (Knowledge engineering techniques)

Copyright 1999, IEE

16/5/37 (Item 4 from file: 94)  
DIALOG(R)File 94:JICST-EPlus  
(c)2004 Japan Science and Tech Corp(JST). All rts. reserv.

04792184 JICST ACCESSION NUMBER: 00A0828090 FILE SEGMENT: JICST-E  
**Interactive Document Retrieval with Relational Learning and it's  
application to the Web Search Engine.**

OKABE MASAYUKI (1); YAMADA SEIJI (1)

(1) Tokyo Inst. of Technology, Graduate School

Denshi Joho Tsushin Gakkai Gijutsu Kenkyu Hokoku(IEIC Technical Report  
(Institute of Electronics, Information and Communication Enginners),  
2000, VOL.100,NO.196(OFS2000 16-22), PAGE.47-52, FIG.2, REF.8

JOURNAL NUMBER: S0532BBG

UNIVERSAL DECIMAL CLASSIFICATION: 681.3:654 002.5:005

LANGUAGE: Japanese COUNTRY OF PUBLICATION: Japan

DOCUMENT TYPE: Journal

ARTICLE TYPE: Original paper

MEDIA TYPE: Printed Publication

ABSTRACT: This paper describes an interactive information **retrieval**  
system which mediates between the user and the web **search** engine to  
realize relevance feedback. This system automatically generates a  
**query** for **search** engine by converting rules which are learned to  
distinguish between **relevant pages** and irrelevant ones user has  
already seen and judged during **retrieval**. The rules consist of  
**keywords**, operators, and region restriction to represent the **feature**  
of the web **page**. They are learned by the algorithm which adopts  
separate-and-conquer strategy and top down heuristic **search** using  
information gain. (author abst.)

DESCRIPTORS: information **retrieval**; WWW(communication); internet; dialog  
processing; learning; feedback; recognition; information system

BROADER DESCRIPTORS: **retrieval**; computer application system; system;  
computer network; communication network; information network; network;  
treatment

CLASSIFICATION CODE(S): JC03000K; AC06020S

Set	Items	Description
S1	6530	AU=(ABE S? OR ABE, S?)
S2	746	S1 AND IC=(G06F? OR H04L?)
S3	44	S2 AND (SEARCH? OR SEEK? OR RETREIV? OR FIND? OR LOCATE? OR LOCATING OR QUER?)
S4	44	IDPAT (sorted in duplicate/non-duplicate order)
S5	41	IDPAT (primary/non-duplicate records only)

? show files

File 344:Chinese Patents Abs Aug 1985-2003/Nov

(c) 2003 European Patent Office

File 347:JAPIO Oct 1976-2003/Sep(Updated 040105)

(c) 2004 JPO & JAPIO

File 348:EUROPEAN PATENTS 1978-2004/Jan W05

(c) 2004 European Patent Office

File 349:PCT FULLTEXT 1979-2002/UB=20040129,UT=20040122

(c) 2004 WIPO/Univentio

File 350:Derwent WPIX 1963-2004/UD,UM &UP=200407

(c) 2004 Thomson Derwent

5/5/1 (Item 1 from file: 350)

DIALOG(R)File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

014228203 \*\*Image available\*\*

WPI Acc No: 2002-048901/200206

XRPX Acc No: N02-036209

**Document information search apparatus for document management system,  
forms keyword from file contents transmitted from search condition  
designating unit, for searching similar documents from database**

Patent Assignee: FUJITSU LTD (FUIT )

Inventor: **ABE S**

Number of Countries: 002 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20010047351	A1	20011129	US 2001761222	A	20010116	200206 B
JP 2002049638	A	20020215	JP 2001131097	A	20010427	200215

Priority Applications (No Type Date): JP 2000155867 A 20000526

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
-----------	------	-----	----	----------	--------------

US 20010047351	A1		22	G06F-007/00	
----------------	----	--	----	-------------	--

JP 2002049638	A		15	G06F-017/30	
---------------	---	--	----	-------------	--

Abstract (Basic): US 20010047351 A1

NOVELTY - A **search** condition designating unit of a **search** unit, designates a file as a **search** condition and transmits content of designated file through network. A document **search** unit (30) forms a keyword, from file contents transmitted from **search** condition designating unit for **searching** similar documents from a database (22).

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(a) Document information **search** method;

(b) Computer readable recorded medium storing document information **search** program

USE - For **searching** documents including interesting contents by e-mail, internet, etc., for management by document management system.

ADVANTAGE - Enables document similar to a document which is not registered in a **search** database to be promptly **searched** by a simple operation.

DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of document **search** unit..

Database (22)

Document **search** unit (30)

pp; 22 DwgNo 3/12

Title Terms: DOCUMENT; INFORMATION; **SEARCH** ; APPARATUS; DOCUMENT;

MANAGEMENT; SYSTEM; FORM; KEYWORD; FILE; CONTENT; TRANSMIT; **SEARCH** ;

CONDITION; DESIGNATED; UNIT; **SEARCH** ; SIMILAR; DOCUMENT; DATABASE

Derwent Class: T01

International Patent Class (Main): G06F-007/00 ; G06F-017/30

File Segment: EPI

5/5/25 (Item 25 from file: 347)

DIALOG(R) File 347:JAPIO

(c) 2004 JPO & JAPIO. All rts. reserv.

07139059 \*\*Image available\*\*

INFORMATION **SEARCHING** DEVICE, INFORMATION **SEARCHING** METHOD AND  
RECORDING MEDIUM

PUB. NO.: 2002-007431 [JP 2002007431 A]

PUBLISHED: January 11, 2002 (20020111)

INVENTOR(s): INOUE SATOSHI

**ABE SHINJI**

TONOMURA YOSHINOBU

APPLICANT(s): NIPPON TELEGR & TELEPH CORP (NTT)

APPL. NO.: 2000-193529 [JP 2000193529]

FILED: June 27, 2000 (20000627)

INTL CLASS: **G06F-017/30 ; G06F-013/00**

#### ABSTRACT

PROBLEM TO BE SOLVED: To provide an information **searching** device, an information **searching** method and a recording medium by which only necessary contents among contents following from a prescribed start page can be acquired by giving a **query** such as 'I wish to acquire a homepage including one or more images, using 'KYOTO' and 'SIGHTSEEING' as a keyword and described in 'Japanese' to a **searching** robot.

SOLUTION: Only the necessary contents are acquired by checking whether or not the contents is compatible to the inputted **query** .

COPYRIGHT: (C)2002,JPO

Set	Items	Description
S1	1369	E3,E4,E6,E7,E8,E9
S2	152	S1 AND (SEARCH? OR SEEK? OR FIND? OR QUER? OR LOCAT? OR RETRIEV?)
S3	0	S2 AND (KEYWORD? OR (THESAURUS OR INDEX OR KEY) () (WORD? OR TERM? OR PHRASE?) OR KEYTERM? OR KEYPHRASE?)

? show files

File 2:INSPEC 1969-2004/Jan W4  
(c) 2004 Institution of Electrical Engineers

File 8:EI Compendex(R) 1970-2004/Jan W3  
(c) 2004 Elsevier Eng. Info. Inc.

File 16:Gale Group PROMT(R) 1990-2004/Feb 03  
(c) 2004 The Gale Group

File 34:SciSearch(R) Cited Ref Sci 1990-2004/Jan W4  
(c) 2004 Inst for Sci Info

File 35:Dissertation Abs Online 1861-2004/Jan  
(c) 2004 ProQuest Info&Learning

File 65:Inside Conferences 1993-2004/Feb W1  
(c) 2004 BLDSC all rts. reserv.

File 148:Gale Group Trade & Industry DB 1976-2004/Feb 03  
(c)2004 The Gale Group

File 647:CMP Computer Fulltext 1988-2004/Jan W4  
(c) 2004 CMP Media, LLC

File 674:Computer News Fulltext 1989-2004/Jan W4  
(c) 2004 IDG Communications

File 275:Gale Group Computer DB(TM) 1983-2004/Feb 03  
(c) 2004 The Gale Group

File 636:Gale Group Newsletter DB(TM) 1987-2004/Feb 03  
(c) 2004 The Gale Group